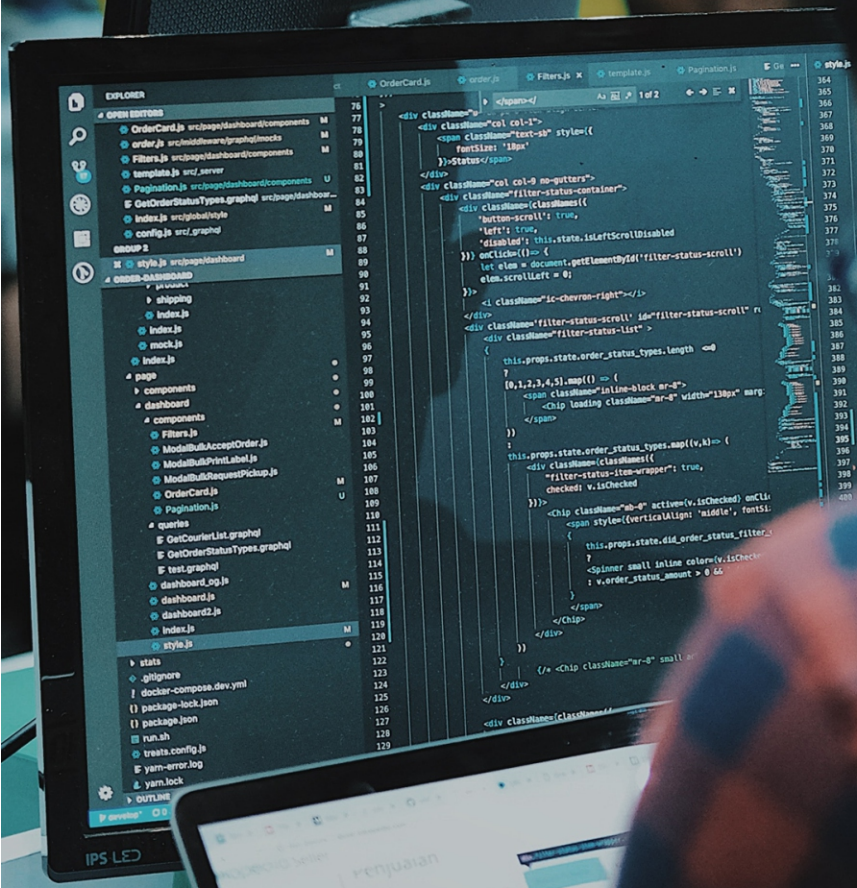


GLOBAL JOURNAL

OF COMPUTER SCIENCE AND TECHNOLOGY: C

Software & Data Engineering



Analyzing Medical Images

Optimize Products, Services

Highlights

A Software Design Pattern

Maintenance of Fertilizer Plants

Discovering Thoughts, Inventing Future



GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY: C
SOFTWARE & DATA ENGINEERING



GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY: C
SOFTWARE & DATA ENGINEERING

VOLUME 22 ISSUE 2 (VER. 1.0)

© Global Journal of Computer Science and Technology. 2022.

All rights reserved.

This is a special issue published in version 1.0 of "Global Journal of Computer Science and Technology" By Global Journals Inc.

All articles are open access articles distributed under "Global Journal of Computer Science and Technology"

Reading License, which permits restricted use. Entire contents are copyright by of "Global Journal of Computer Science and Technology" unless otherwise noted on specific articles.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without written permission.

The opinions and statements made in this book are those of the authors concerned. Ultraculture has not verified and neither confirms nor denies any of the foregoing and no warranty or fitness is implied.

Engage with the contents herein at your own risk.

The use of this journal, and the terms and conditions for our providing information, is governed by our Disclaimer, Terms and Conditions and Privacy Policy given on our website <http://globaljournals.us/terms-and-condition/menu-id-1463/>

By referring / using / reading / any type of association / referencing this journal, this signifies and you acknowledge that you have read them and that you accept and will be bound by the terms thereof.

All information, journals, this journal, activities undertaken, materials, services and our website, terms and conditions, privacy policy, and this journal is subject to change anytime without any prior notice.

Incorporation No.: 0423089
License No.: 42125/022010/1186
Registration No.: 430374
Import-Export Code: 1109007027
Employer Identification Number (EIN):
USA Tax ID: 98-0673427

Global Journals Inc.

(A Delaware USA Incorporation with "Good Standing"; Reg. Number: 0423089)

Sponsors: Open Association of Research Society

Open Scientific Standards

Publisher's Headquarters office

Global Journals® Headquarters
945th Concord Streets,
Framingham Massachusetts Pin: 01701,
United States of America

USA Toll Free: +001-888-839-7392

USA Toll Free Fax: +001-888-839-7392

Offset Typesetting

Global Journals Incorporated
2nd, Lansdowne, Lansdowne Rd., Croydon-Surrey,
Pin: CR9 2ER, United Kingdom

Packaging & Continental Dispatching

Global Journals Pvt Ltd
E-3130 Sudama Nagar, Near Gopur Square,
Indore, M.P., Pin:452009, India

Find a correspondence nodal officer near you

To find nodal officer of your country, please
email us at local@globaljournals.org

eContacts

Press Inquiries: press@globaljournals.org
Investor Inquiries: investors@globaljournals.org
Technical Support: technology@globaljournals.org
Media & Releases: media@globaljournals.org

Pricing (Excluding Air Parcel Charges):

Yearly Subscription (Personal & Institutional)
250 USD (B/W) & 350 USD (Color)

EDITORIAL BOARD

GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY

Dr. Corina Sas

School of Computing and Communication
Lancaster University Lancaster, UK

Dr. Sotiris Kotsiantis

Ph.D. in Computer Science, Department of Mathematics,
University of Patras, Greece

Dr. Diego Gonzalez-Aguilera

Ph.D. in Photogrammetry and Computer Vision Head of
the Cartographic and Land Engineering Department
University of Salamanca Spain

Dr. Yuanyang Zhang

Ph.D. of Computer Science, B.S. of Electrical and
Computer Engineering, University of California, Santa
Barbara, United States

Dr. Osman Balci, Professor

Department of Computer Science Virginia Tech, Virginia
University Ph.D. and M.S. Syracuse University, Syracuse,
New York M.S. and B.S. Bogazici University, Istanbul,
Turkey

Dr. Kwan Min Lee

Ph. D., Communication, MA, Telecommunication,
Nanyang Technological University, Singapore

Dr. Khalid Nazim Abdul Sattar

Ph.D, B.E., M.Tech, MBA, Majmaah University,
Saudi Arabia

Dr. Jianyuan Min

Ph.D. in Computer Science, M.S. in Computer Science, B.S.
in Computer Science, Texas A&M University, United States

Dr. Kassim Mwitondi

M.Sc., PGCLT, Ph.D. Senior Lecturer Applied Statistics/
Data Mining, Sheffield Hallam University, UK

Dr. Kurt Maly

Ph.D. in Computer Networks, New York University,
Department of Computer Science Old Dominion
University, Norfolk, Virginia

Dr. Zhengyu Yang

Ph.D. in Computer Engineering, M.Sc. in
Telecommunications, B.Sc. in Communication Engineering,
Northeastern University, Boston, United States

Dr. Don. S

Ph.D in Computer, Information and Communication
Engineering, M.Tech in Computer Cognition Technology,
B.Sc in Computer Science, Konkuk University, South
Korea

Dr. Ramadan Elaiess

Ph.D in Computer and Information Science, University of
Benghazi, Libya

Dr. Omar Ahmed Abed Alzubi

Ph.D in Computer and Network Security, Al-Balqa Applied
University, Jordan

Dr. Stefano Berretti

Ph.D. in Computer Engineering and Telecommunications, University of Firenze Professor Department of Information Engineering, University of Firenze, Italy

Dr. Lamri Sayad

Ph.d in Computer science, University of BEJAIA, Algeria

Dr. Hazra Imran

Ph.D in Computer Science (Information Retrieval), Athabasca University, Canada

Dr. Nurul Akmar Binti Emran

Ph.D in Computer Science, MSc in Computer Science, Universiti Teknikal Malaysia Melaka, Malaysia

Dr. Anis Bey

Dept. of Computer Science, Badji Mokhtar-Annaba University, Annaba, Algeria

Dr. Rajesh Kumar Rolan

Ph.D in Computer Science, MCA & BCA - IGNOU, MCTS & MCP - Microsoft, SCJP - Sun Microsystems, Singhania University, India

Dr. Aziz M. Barbar

Ph.D. IEEE Senior Member Chairperson, Department of Computer Science AUST - American University of Science & Technology Alfred Naccash Avenue Ashrafieh, Lebanon

Dr. Chutisant Kerdvibulvech

Dept. of Inf. & Commun. Technol., Rangsit University Pathum Thani, Thailand Chulalongkorn University Ph.D. Thailand Keio University, Tokyo, Japan

Dr. Abdurrahman Arslanyilmaz

Computer Science & Information Systems Department Youngstown State University Ph.D., Texas A&M University University of Missouri, Columbia Gazi University, Turkey

Dr. Tauqeer Ahmad Usmani

Ph.D in Computer Science, Oman

Dr. Magdy Shayboub Ali

Ph.D in Computer Sciences, MSc in Computer Sciences and Engineering, BSc in Electronic Engineering, Suez Canal University, Egypt

Dr. Asim Sinan Yuksel

Ph.D in Computer Engineering, M.Sc., B.Eng., Suleyman Demirel University, Turkey

Alessandra Lumini

Associate Researcher Department of Computer Science and Engineering University of Bologna Italy

Dr. Rajneesh Kumar Gujral

Ph.D in Computer Science and Engineering, M.TECH in Information Technology, B. E. in Computer Science and Engineering, CCNA Certified Network Instructor, Diploma Course in Computer Servicing and Maintenance (DCS), Maharishi Markandeshwar University Mullana, India

Dr. Federico Tramarin

Ph.D., Computer Engineering and Networks Group, Institute of Electronics, Italy Department of Information Engineering of the University of Padova, Italy

Dr. Roheet Bhatnagar

Ph.D in Computer Science, B.Tech in Computer Science, M.Tech in Remote Sensing, Sikkim Manipal University, India

CONTENTS OF THE ISSUE

- i. Copyright Notice
 - ii. Editorial Board Members
 - iii. Chief Author and Dean
 - iv. Contents of the Issue
-
- 1. Fake News Detection: Covid-19 Perspective. *1-12*
 - 2. Capillary X: A Software Design Pattern for Analyzing Medical Images in Real-time using Deep Learning. *13-23*
 - 3. Steam Package Boiler Expert System for Control and Maintenance of Fertilizer Plants using Rule-Base Fuzzy Logic. *25-39*
 - 4. Integration of the Big Data Environment in a Financial Sector Entity to Optimize Products, Services and Decision-Making. *41-51*
-
- v. Fellows
 - vi. Auxiliary Memberships
 - vii. Preferred Author Guidelines
 - viii. Index



GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY: C
SOFTWARE & DATA ENGINEERING
Volume 22 Issue 2 Version 1.0 Year 2022
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Fake News Detection: Covid-19 Perspective

By Md. Ziaur Rahman Shamim, Shaheena Sultana, Anika Tabassum,
Israt Tabassum & Sarkar Binoyee Farha

Notre Dame University

Abstract- The development of social media has contributed to a remarkable rise in the spread of fake news. Today people rely more on online news outlets. The chance of receiving fake news on an online platform is high. As we went through a pandemic and the Covid-19 was the most absorbing topic of 2020, much news on Covid-19 was published every day in traditional media and social media. Among that news, some are fake. In this work, we have collected a new dataset for detecting fake news from traditional media on Covid-19. We have gathered more than 3000 pieces of news from traditional media out of the 170 are fake ones that were collected from fact-checking sites. Then we have tested the existing four classification algorithms with our dataset using Count Vectorizer and TF-IDF. We have merged 170 fake news with four scales of true news and analyzed the outcome.

Index Terms: fake news, fake news detection, traditional media, data set, covid-19, social media.

GJCST-C Classification: DDC Code: 004.678 LCC Code: TK5105.875.I57



Strictly as per the compliance and regulations of:



© 2022. Md. Ziaur Rahman Shamim, Shaheena Sultana, Anika Tabassum, Israt Tabassum & Sarkar Binoyee Farha. This research/review article is distributed under the terms of the Attribution-NonCommercial-NoDerivatives 4.0 International (CC BYNCND 4.0). You must give appropriate credit to authors and reference this article if parts of the article are reproduced in any manner. Applicable licensing terms are at <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

Fake News Detection: Covid-19 Perspective

Md. Ziaur Rahman Shamim ^α, Shaheena Sultana ^σ, Anika Tabassum ^ρ, Israt Tabassum ^ω
& Sarkar Binoyee Farha [¥]

Abstract- The development of social media has contributed to a remarkable rise in the spread of fake news. Today people rely more on online news outlets. The chance of receiving fake news on an online platform is high. As we went through a pandemic and the Covid-19 was the most absorbing topic of 2020, much news on Covid-19 was published every day in traditional media and social media. Among that news, some are fake. In this work, we have collected a new dataset for detecting fake news from traditional media on Covid-19. We have gathered more than 3000 pieces of news from traditional media out of the 170 are fake ones that were collected from fact-checking sites. Then we have tested the existing four classification algorithms with our dataset using Count Vectorizer and TF-IDF. We have merged 170 fake news with four scales of true news and analyzed the outcome.

Index Terms: fake news, fake news detection, traditional media, data set, covid-19, social media.

I. INTRODUCTION

In 2020, we have gone through trauma due to covid-19. This pandemic is not less than a trauma. Covid-19 has become a challenge to each and everyone whose country is facing a huge number of positive cases regarding this and the number of deaths are rising on a daily basis, which actually makes one depressed. During this pandemic situation people are stuck in their own places thus they are consuming much more internet than before. According to openvault, data usage has increased 47 percent during this pandemic [34]. As people are consuming more internet, they are also getting their news from it. Among those news stories, some of them are fake, but it is hard for maximum people to find out whether the news is fake or not. Through an automated system we can easily detect fake news. An automated system could be defined as a technique that performs a task using programming inputs and computerized feedback control to verify that the instructions are followed correctly. The end result is a technology that can function without any need for human interaction [18]. An automated system needs a well-labeled data set to detect fake news from tons of news. So, we have made a new data set on Covid-19.

Author ^α ^σ ^ρ ^ω [¥]: Department of Computer Science and Engineering, Notre Dame University Bangladesh Dhaka, Bangladesh.
e-mails: zrshamim8822@gmail.com, shaheenacse@ndub.edu.bd, anika4725@gmail.com, israt.tabassum34@gmail.com, farha1binoyee@gmail.com

What is Fake News? Fake news is misleading or false news information introduced as news [19]. There is no valid definition of fake news. Many researchers [32], [11] have adopted Hunt Allcot et al. definition. They defined, a news article that is deliberately and verifiably false and could deceive readers is called fake news [4]. Also another explanation from S. Desai, fake news is those articles that are fabricated, false information with no supportable facts, origin, or quotes [13]. These two definitions have two key features: authenticity, fake news includes false information, facts, sources, or quotes to make it more authentic so that readers can have no doubt about the news and another is intent. Basically, fake news is generated with the purpose of deceiving readers, sometimes to harm individuals. In some papers, satire or parody news are also considered as fake news [7], [8]. Fake news can be spread via two types of media such as traditional media and social media [32]. Traditional media are newspapers, magazines, blogs, etc. and social media are Facebook, Twitter, WhatsApp, etc. So basically, we have focused our data set to detect fake news from traditional media and news related to Covid-19.

According to researchers, in the first three months of the year 2020 at least 800 people might have died across the world because of fake news about Covid-19 [10]. Fake news is not only spreading disinformation or rumors, but also taking valuable lives and also hampering public properties. To check the facts of fake news there are some delegated websites like Politifact [20], lead stories [15], FactCheck [28], the full fact [24], health feedback [35] where they manually updates potential fake news stories which are published online platforms.

We have collected a vast number of news data sets so that our outcome can help the upcoming researchers to work on this problem. We have worked on Covid-19 which is the most hooked and challenging topic in the world and every day we have got thousands of news regarding this pandemic. From those news some of them are fake. In the beginning, people used to believe in every news regarding Covid-19. By these, some of them became so panic that they were about to lose their lives and some of them already lost theirs. There were too many rumors that people started believing those. For example, in Iran, there were rumors across the country that alcohol consumption would help combat Covid-19 [16], [5].

Hence, people started consuming alcohol in large numbers. Like this news, rumors were being spread across the world and people started to believe that news. In this work, we have presented a new data set to detect fake news. For this data set, we have gathered more than three thousand news from various news sites. In our data set, there are two types of news, one is true news another one is fake news. There are one hundred and seventy fake news in our data set. For detecting fake news from our data set, we have used four classification algorithm they are multinomial Naïve bayes, logistic regression, support vector machine and passive aggressive classifier.

The rest of the work is organized as follows: in chapter II, we have reviewed some of the previous work. In chapter III, we have discussed the algorithm that we have used for our work. In chapter IV, we have discussed our work. In chapter V, we have discussed the performance and have concluded our work in chapter VI.

II. LITERATURE REVIEW

Previously many researchers have worked on this particular area. Some of them created new models or explore new areas to detect fake news. Others improved the existing model to improve fake news detection and others build a robust data set to fuel the detection system. In this section, we explore some of the previous works. Ruchansky et al. presented a fake news detection System using a Hybrid Deep Model. They proposed a model called the CSI. Capture, Score, and Integrate are the three modules of CSI model. The capture module is based on text and reaction. This module captured the engagement between users and news articles, the second module score learns the source feature. Then they combined two modules and integrated them with the third module to produce a label for fake news [29]. Kai Shu et al. presented a brief analysis of detecting fake news from social media, they also discussed fake News Classification based on social context and psychology, also reviewed current algorithm in terms of data mining [32]. Lutzke et al. conducted a 3 by 2 experimental model having guidelines, enhanced guidelines, and a controlled part. Each experiment has two types of news about climate change one is fake and another is real. A total of 2750 people from different fields participated in their experiment. Each participant was randomly allocate to one of six possible experimental variations [22]. Bahad et al. used deep learning model, Bi-directional Long short-term memory (LSTM) over other techniques like as Convolutional Neural Networks (CNN), unidirectional LSTM, Recurrent neural network (RNN), for detecting fake news [6]. Abdullah et al. used machine learning

and deep learning technique to detect fake news from Twitter data set. They have used five different Machine Learning algorithms, like Support Vector Machine, Naïve - Bayes algorithm, Logistic Regression, and Recurrent Neural Network to detect fake news from the data set [1]. Zhang et al. used a two-layer method that includes the identification of fake topics and fake incidents [38]. Yang et al. detect satirical news using linguistic features [37]. Apuke and Omar developed a complete model studying from uses and gratification view [14]. Agarwal et al. used the natural language processing and the machine learning method to detect fake news [2]. Granik and Mesyura used a Naïve- Bayes classifier to classify news from BuzzFeed data sets [17]. Yang presented Liar data set containing news more than 12 thousand for fake news detection. The data set is divided into six classes [36]. Fake News Corpus by Szpakowski. He uses multiple corpora to develop and test various models. The corpus has been automatically crawled using open sources.co labels [33]. Khan et al. performed benchmark research to compare the performance of several machine learning methods on three distinct data sets [21]. Zhang and Ghorbani gave a detailed review of what has been discovered so far on false news. They have described the detrimental impact of online fake news as well as the current status of detecting tools [39]. Alkhodair et al. presented a novel method for automatically identifying rumors, combining the learning of word embedding with the training of a recurrent neural network with two separate goals [3].

III. ALGORITHMS

Previously many algorithms have been developed to detect fake news. We have used four different types of supervised learning algorithms to detect fake news. We briefly discuss the four existing detection algorithms in this section.

In machine learning, Supervised learning is the task of creating a function that maps an input to an output based on given input-output pairs [30]. In supervised learning, the program is given labeled input data and the output result is expected. In regression and classification problems, supervised learning works fine, such as deciding what group a news story belongs to.

a) *Multinomial Naïve Bayes*

Naïve bayes classifiers are members of "probabilistic classifiers" which is build from applied bayes theorem with strong (Naïve) independent assumptions between the attributes. Due to its simplicity, speed, and good accuracy Multinomial Naïve bayes [23] is a common classifier for text classification. Multinomial classifier learns a conditional probability that the nth document x^n form a

class y_c given the document x^n that is formed by Bayes rule, $p(y_c|x^j) = \frac{p(y_c)p(x^j|y_c)}{p(x^j)}$, Multinomial Naïve bayes uses multinomial model to find out $x^j p(f_i|y_c)$. So, in particular

$p(x^j|y_c) = \pi_{i=1}^{n_f} p(f_i|y_c)^{x_i^j} n_f$, is the number of feature x_i^j is the number of times the i th feature occurs in the j th document x^j $p(f_i|y_c)$ is the probability of the i th feature occurring given class y_c

In text classification, the conditional feature probabilities are estimated using Laplace smoothing. Laplace smoothing can be written as,

$$P_{Laplace}(f_i|y_c) = \frac{\lambda + \sum_{j=1}^{n_d} x_i^j}{\sum_{k=1}^{n_f} [\lambda + \sum_{j'}^{n_d} x_{j'}^k]}$$

Here, n_d is the number of data points in the training set from class y_c n_f , is the number of features λ , is a parameter known as the Laplace smoothing constant. λ is typically set to 1. As λ is a non-zero number, it prohibits such degenerate cases from equaling zero. In general, if a feature does not occur in any document of the training set, $p(x^j|y_c)$ or all documents x_j . The test set that contains the degenerated feature will be zero for all classes y_c , causing Multinomial Naïve bayes to lose all discriminating power.

b) *Logistic Regression*

Logistic regression is a classification model that models a binary variable using a logistic function [31]. It is a categorical dependent variable prediction approach that uses a collection of independent variables to forecast a dependent variable.

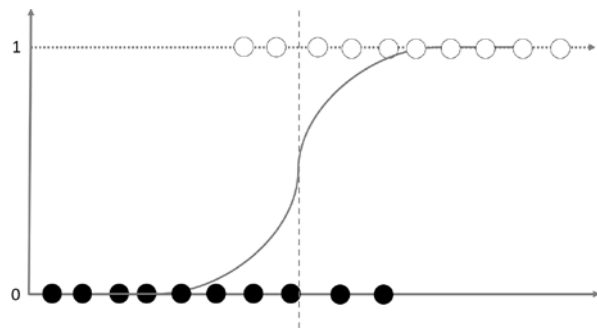


Fig.1: Example of a Logistic Regression

Because of the application of the sigmoid function in logistic regression, the curve in Fig. 1 is generated. The sigmoid curve is another name for the curve produced above.

c) *Support Vector Machine*

Vapnik and Cortes proposed a Support vector machine (SVM). The Support vector machine is a supervised machine learning method used for classification, regression problems. It determines the best decision boundary among various vectors [12]. Support Vector Machine draws a hyper-plane to separate two distinct classes [25].

Support Vector Machines not only create hyper-plane but also construct a maximum margin separator, a decision boundary, with the largest possible distance. SVM constructs the marginal line Fig. 2 by the nearest one or more positive or negative points, these points are called to be support vectors. The Marginal lines are drawn

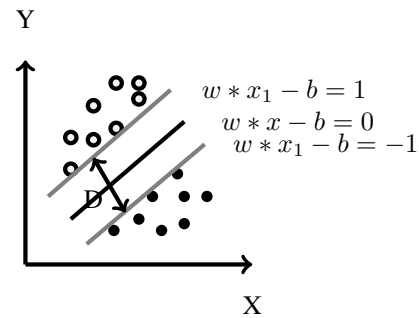


Fig. 2: Support Vector Machine

parallel with the hyper-plane and the distance between two marginal lines with the hyper-plane is equal. The total distance (D) between the two margins is called to be marginal distance. For having a better model SVM finds the largest distance (D) between support vectors in shorts marginal lines.

SVM produce a linear hyper-line of separation between classes which are linearly separable. But in nonlinear classification problems like as Fig. 3, it is not possible to have better classification only by drawing a hyper-plane. To solve Fig. 3a this kind of problem support vector machine using the so-called kernel trick.

They have the capability to implant the data from a lower dimension into higher-dimensional space. In Fig. 3b we can see that, the kernel trick takes the lower space input points Fig. 3a and then implant them into a higher dimension.

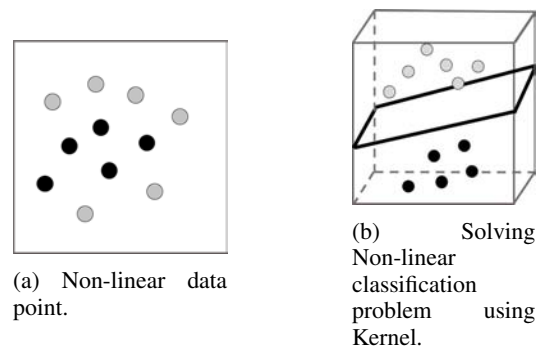


Fig. 3: A Non-linear classification problem.

d) *Passive Aggressive Classifier*

The passive-aggressive algorithm is a member of Machine learning algorithms. Passive-Aggressive algorithms are called so because :

- *Passive*: Keep the model and do not make any adjustments if the prediction is accurate.
- *Aggressive*: Make adjustments to the model if the prediction is incorrect.

The key idea of passive-aggressive classifier is that with each misclassified training data point it gets, the classifier changes its weight vector, attempting to fix it.

Initially, the weight W is 0, then the algorithm receives a new doc, then the algorithm applies feature extraction and normalizes the doc. If the $d^T w$ value is greater than 0 the algorithm will predict the class as a positive and prediction class will be $y = \pm 1$.

Algorithm 1 Passive Aggressive Classifier.

Initialize $w = (0, \dots, 0)$

Receive a new doc $d = d_1, \dots, d_v$

Apply feature extraction, normalize $\|d\| = 1$

Predict positive if $d^T w > 0$

Observe true class : $y = \pm 1$

Want to have:

$d^T w \geq +1$ if positive ($y = +1$)

$d^T w \leq -1$ if negative ($y = -1$)

Same as: $y d^T w \geq 1$

Loss : $L = \max(0, 1 - y(d^T w))$

Update: $w_{new} = w + yLd$

The core concept of the passive-aggressive classifier is to determine the loss function and act according to it. So, the loss function of the passive-aggressive classifier is Loss: $L = \max(0, 1 - y(d^T w))$. If there any misclassification, there will be a loss. If the data set has an input class that belongs to a positive class, but initially the output is a negative class. For this misclassification, the passive-aggressive classifier will update its initial weight.

IV. OUR WORK

A workflow is a sequence of tasks that processes a set of works. In Fig. 4 we have shown our workflow. Firstly, we have collected true and fake news from the various traditional media. As our data are text data so for feature extraction we approach through two different feature extractors one is TF-IDF (Term Frequency and Inverse Document Frequency) and another is count vectorizer. Then we have tested them with four different supervised algorithms and finally discussed the final result based on our data set.

a) Data Collection Process

The main focus of our work is to collect news from various traditional news media as shown in Fig.

6. We choose the most popular and reputable news websites from across the world to construct a collection of authentic news. Such as Aljazeera, The New York Times, BBC, CNN, and many other news websites as well. For fake news, we have collected from some reputed fact-checking websites such as Lead Stories, PolitiFact, Health Feedback, etc. These fact-checking websites offer a clear and instructive explanation of fake news that has previously been published on other websites. We have the nine metadata from news sites, the title of a news annotated as a headline, the main article referred to the body, URL of the news, the date of publication, publisher name, the type of the article. In the fake news data set, we added two additional attributes, one is a fact-checking website name and another one is URLs of fact articles for the authentication of the news. An example of the data collection process for our data set is shown in Fig. 5. Our collected data set is given in the link 1. We have labeled the true as 1 and the fake news as 0.

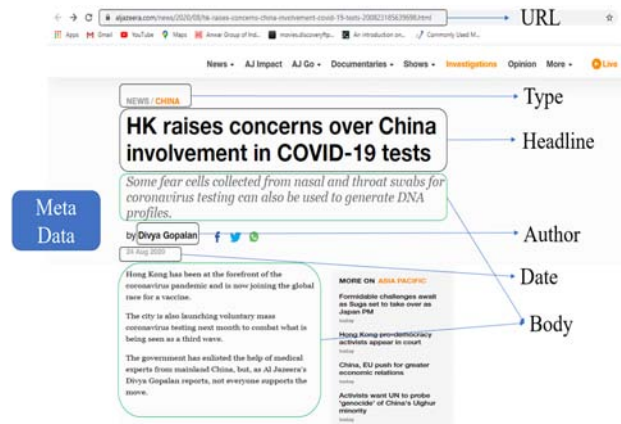


Fig. 5: Meta data example

b) Data Preprocessing

Data preprocessing is a data mining approach that includes transforming raw data into a suitable state. Raw data is frequently incomplete, inaccurate, and even absent, and it is prone to have many faults in certain actions or patterns. Preprocessing data is a well-known approach for resolving such issues.

Initially, input data is typically provided in the raw state, which is in the texts, sentences, comments, articles, messages etc format. It requires certain clean-up or preprocessing before the data can be moved to machine learning algorithms, so algorithms can focus on key terms instead of terms that adds limited or no importance. So in raw state text data, there is a special character. Non-alphanumeric characters, as we all know, are special characters, these symbols are most frequently found in text data. These symbols contribute little meaning to text

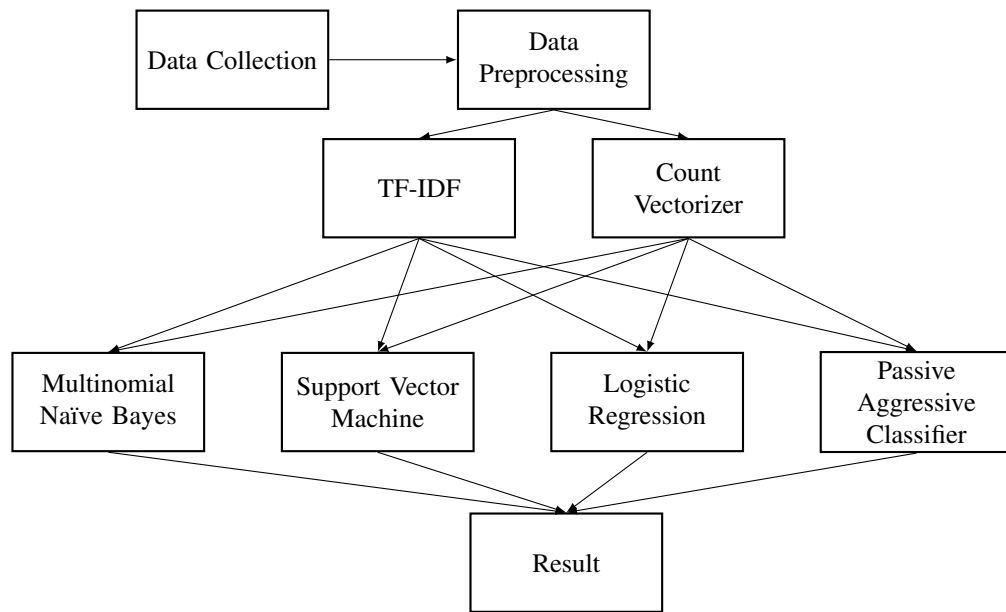


Fig. 4: Workflow of fake news detection: the Covid-19 perspective

noise. So, to avoid noise the whole raw data need to convert into a single form either all words into small character or capital character. For the further addition we have converted our data into a set of arrays to perform stemming and removing stop words.

c) Porter Stemmer

For our work, we have used porter stemmer for stemming. In 1980 a stemming method has gained much popularity which is now known as porter Stammer [26]. Speed and simplicity, are its identification. Data mining and Information retrieval are the main application of porter stammer. As porter stammer uses suffix stripping to produce stems. It produces the best output as compared to other stammers and also it has less error rate than other techniques. However, English words are limited to their application. The output of stem might not be a meaningful word, but the collection of stems is plotted onto the same stem as well.

d) Stop Words

There are certain words in a sentence as an example, be, too, not, etc. in the English language, which does not have any significance for the processing of natural language. So, during natural language processing, such words are taken out. In fact, stop words are some words that are stripped out of the processing of natural language [27]. In particular, stop words do not bring a lot of value to natural language processing results. Without modifying or compromising the context of every tatement, we can comfortably neglect it.

e) Feature Extraction

As our data is text data containing words. So, for use as inputs in machine learning algorithms, these terms now have to be represented as integers, or floatingpoint values. This method is considered the extraction of features or vectorization.

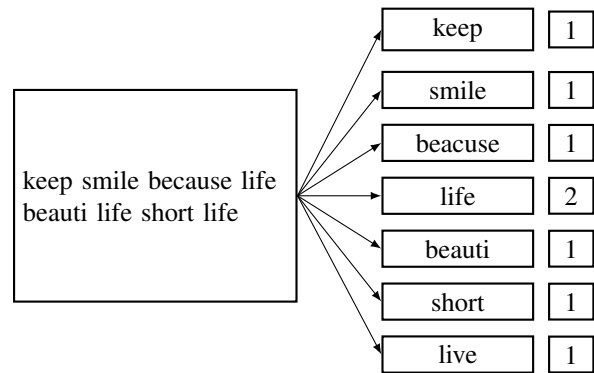


Fig. 8: Count Vectorizer

Count Vectorizer: The count vectorizer is often used to transform a text data collection to a word count vector. This transforms a set of textual data into a token count matrix. As we can see in Fig. 8 count vectorizer converts each word from text data into a count vector. Actually, the count vectorizer counts how frequently a

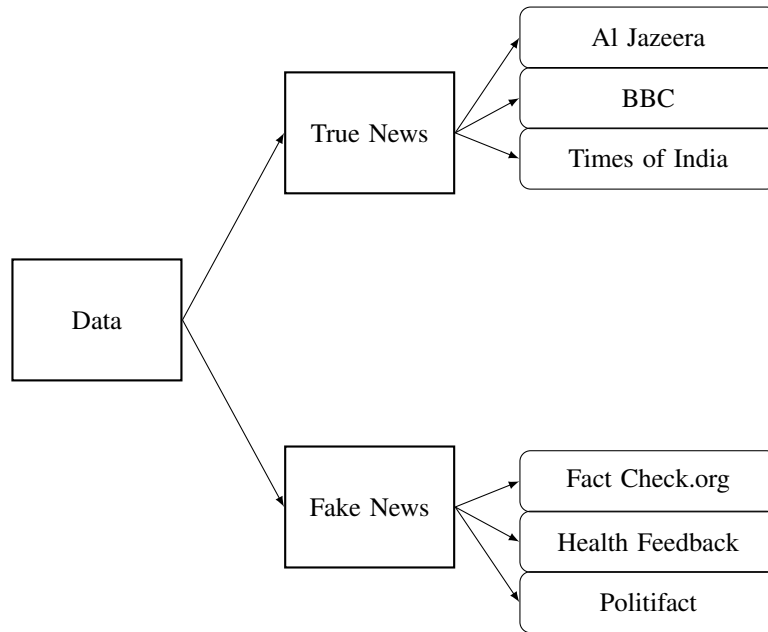


Fig. 6: Data collection process

- *Term Frequency(TF)*: Term frequency calculates the frequency of a term or word occurs in a data set.

$$TF(t) = \frac{\text{Number of times term } t \text{ appears in a document}}{\text{Total number of terms in the document}}$$

- *Inverse document frequency(IDF)*: Inverse document frequency calculates how important a word or term is in a document. In TF all words or terms are equally considered important. Thus, IDF weight down repeated words or terms and scale up the uncommon ones.

$$IDF(t) = \log_e \frac{\text{Total number of documents}}{\text{Number of documents with term } t \text{ in it}}$$

f) *Confusion Matrix*

A confusion matrix is a method of describing the classification algorithms performances [9]. A confusion matrix is a table that shows how well a classification model (or "classifier") performs on a testing data set for which the real values are known. The confusion matrix itself is uncomplicated, but the related terms might be complex.

PREDICTIVE VALUES

POSITIVE (1) NEGATIVE (0)

ACTUAL VALUES

POSITIVE (1)

NEGATIVE (0)

	POSITIVE (1)	NEGATIVE (0)
POSITIVE (1)	TP	FN
NEGATIVE (0)	FP	TN

Fig. 9: Confusion Matrix

- *True Positive(TP)*: When the positive type is predicted exactly by a model.
- *True Negative(TN)*: When the negative type is predicted exactly by a model.
- *False Positive(FP)*: When the positive type is predicted incorrectly by a model.
- *False Negative(FN)*: When the negative type is predicted incorrectly by a model.
- *Accuracy*: Accuracy is the ratio of accurate forecasts to overall forecasts made.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- *Precision*: Precision is the ratio of positive results

- *Precision*: Precision is the ratio of positive results correctly predicted to all positive results predicted.

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

- *Recall*: Recall is the ratio of rightly predicted positive observation of all outcomes in actual class-yes.

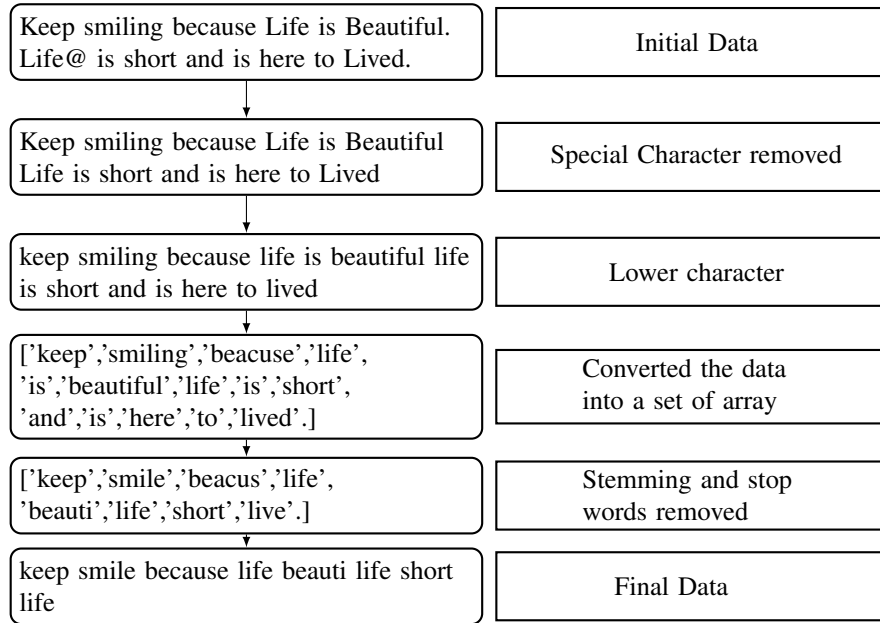


Fig. 7: Data Preprocessing

- *F1-Score*: F1-Score is the weighted average of Precision and Recall.

$$\text{F1-Score} = \frac{2 * \text{Pre} * \text{Re}}{\text{Pre} + \text{Re}}$$

V. PERFORMANCE EVOLUTION

We tested the detection system using two different types of feature extraction one is a count vectorizer and another is TF-IDF. As we have only one hundred seventy fake news so we merged them with four different scales (Table 1) of true data to have a depth look at the detection system.

Table 1: Mixing Fake Data with Four Different Scales of True Data.

	Fake News	True News	Total
Set-1	170	180	350
Set-2	170	230	400
Set-3	170	330	500
Set-4	170	530	800

In set-1 we have merged 170 fake news with 180 true news. In set-2 we have 80 more true news than fake news, 58% of total data is true news. In set-3, 66% of data is truly labeled, and set-4 having 76% of true

news. Then we have divided our data set into the training set and test set.

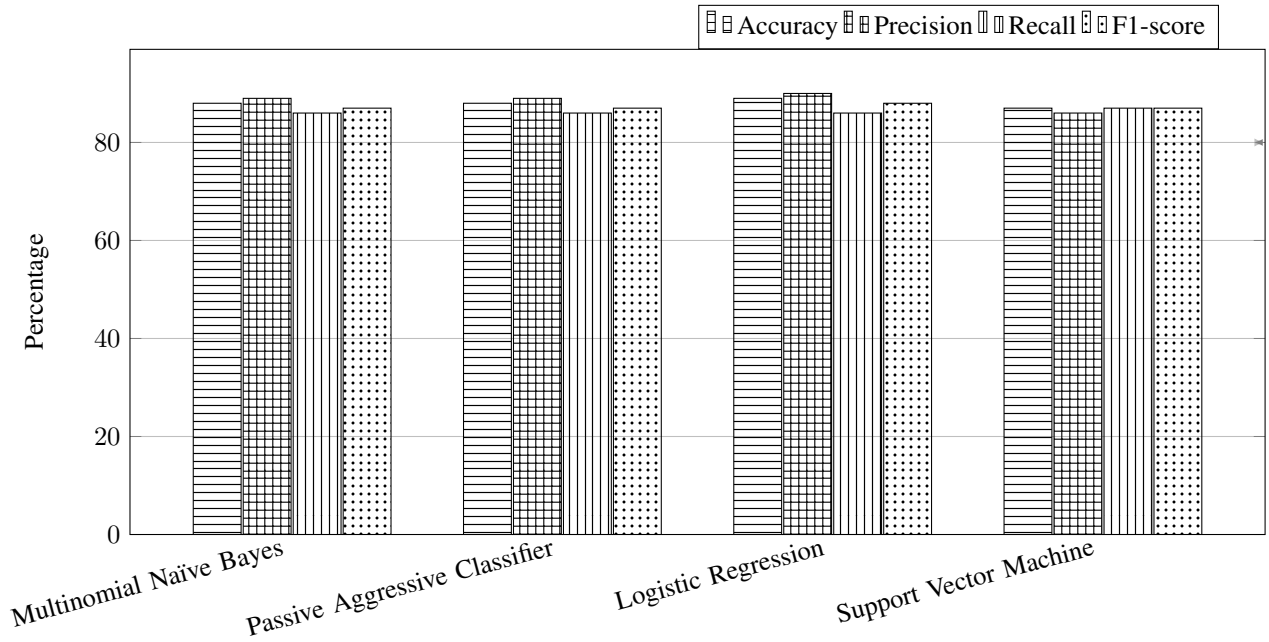
First, we have Multinomial Naïve bayes using TFIDF. We can see our final result in Table 2. Multinomial Naïve bayes shows the highest test accuracy of 89% on set-1 than other three algorithms using TF-IDF. Also shows stable values on precision, recall, and f1-score in fake news detection. In other sets, the test accuracy of Multinomial Naïve bayes declined as the amount of true data is much greater than the fake data. In set 4 it seems that the accuracy of 81% is much better than the previous two sets. But we can see that the recall (8%) and f1-score (14%), which is too much low in detecting fake news. In Table 3, we have our final result of Multinomial Naïve bayes using count vectorizer. We have a different scenario than TF-IDF while using a count vectorizer. In Table 3, we can see that the recall and f1-score drastically decreased while using TF-IDF. But in the count vectorizer, the accuracy increased as expected but the other three values did not drastically decrease shown stable value 3. In set 2 and set 3 Multinomial Naïve bayes shown the same accuracy of 88% (Table 3).

On set 4 Passive aggressive classifier shows highest test accuracy of 92% using TF-IDF than other

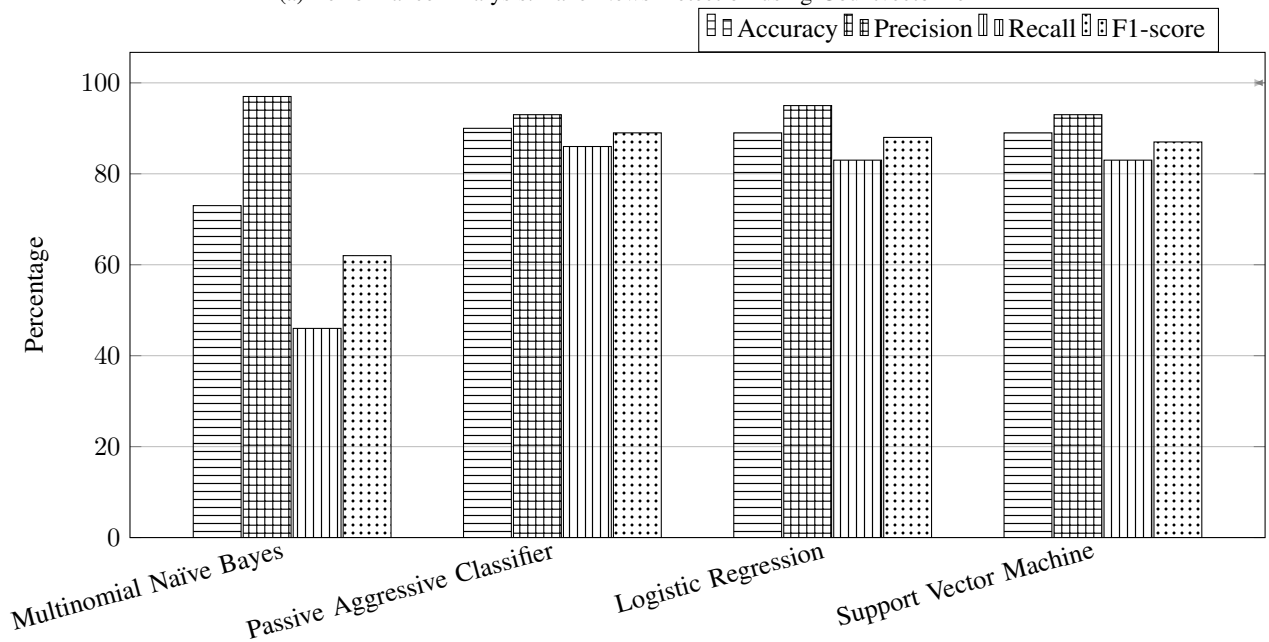
three sets. But the poor value in recall (64%) and f1-score (76%). But in set 2 and set 3 it gives better results on all sections. Passive aggressive classifier using count vectorizer on Table 3. In this part set 2 and set 4 both shown the highest accuracy of 90% then other two sets. But set 4 has a poor outcome on recall (68%) and f1-score (73%) while detecting fake news than other sets, that we have tested. The other three set shown decent outcomes on all three parameters.

Logistic regression also given the highest accuracy of 91% in set 4. But not a satisfying outcome

on the other three-parameter in detecting fake news. In general, set 2 given the best performance out of other four sets, having an accuracy of 89%. In set-4, Fake news detection



(a) Performance Analysis: Fake News Detection using Countvectorizer



(b) Performance Analysis: Fake News Detection using TF-IDF

Fig. 10: Fake News Detection Comparison Between Four Algorithm Using Set-2

Table 2: Fake News Detection Result Using TF-IDF

Technique	Algorithms		T/F	Support	Precision	Recall	F1-Score	Accuracy	
TF-IDF	Multinomial naïve bias	Set-1	0	53	0.93	0.81	0.87	0.89	
			1	63	0.86	0.95	0.9		
		Set-2	0	63	0.97	0.46	0.62	0.73	
			1	69	0.67	0.99	0.8		
		Set-3	0	61	1	0.23	0.37	0.72	
			1	104	0.69	1	0.82		
		Set-4	0	53	1	0.08	0.14	0.81	
			1	211	0.81	1	0.9		
		Passive Aggressive	Set-1	0	53	0.87	0.87	0.87	0.88
				1	63	0.89	0.89	0.89	
			Set-2	0	63	0.93	0.86	0.89	0.9
				1	69	0.88	0.94	0.91	
	Set-3		0	61	0.9	0.77	0.83	0.88	
			1	104	0.88	0.95	0.91		
	Set-4		0	53	0.92	0.64	0.76	0.92	
			1	211	0.92	0.99	0.95		
	Support Vector Machine		Set-1	0	53	0.84	0.87	0.85	0.86
				1	63	0.89	0.86	0.87	
			Set-2	0	63	0.93	0.83	0.87	0.89
				1	69	0.86	0.94	0.9	
		Set-3	0	61	0.96	0.74	0.83	0.89	
			1	104	0.86	0.98	0.92		
		Set-4	0	53	0.89	0.62	0.73	0.91	
			1	211	0.91	0.98	0.95		
Logistic Regression		Set-1	0	53	0.84	0.87	0.85	0.86	
			1	63	0.89	0.86	0.87		
		Set-2	0	63	0.95	0.83	0.88	0.89	
			1	69	0.86	0.96	0.90		
	Set-3	0	61	0.90	0.77	0.83	0.88		
		1	104	0.88	0.95	0.91			
	Set-4	0	53	0.91	0.6	0.73	0.91		
		1	211	0.91	0.99	0.95			

with logistic regression using count vectorizer also shown best performance on accuracy as we can see in Table 3 but the poor result on other three-parameter. The next highest accuracy of 89% is given by set 2 and set 3. Where set 2 given higher precision (90%), recall (86%), and f1-score (88%) at detecting fake news. In last we have our final result of support vector machine using TF-IDF given in Table 2. In here also set 4 shown the highest accuracy (91%). But poor performance on the other three-parameter. Set 2 and set 3 shown better performance as we can see in Table 2. Set 2 shown a precision of 93%, recall of 83%, f1-score of 87%, and accuracy of 89% which is combined best the other three sets. In Table 3, support vector machine using count vectorizer, as usual like other three algorithm support vector machine has shown highest accuracy of 90% in set 4 but the same poor result on other three-parameter. Set 3 shown the second highest accuracy of 88% and 84% on recall, precision, and f1-score. Set 2 gives an accuracy of 87%, the third-highest among four algorithms but having the higher value of precision (86%), recall (87%), and f1-score (87%) the other three algorithms.

Using set 4, almost all algorithms have given the highest accuracy except Multinomial Naïve bayes

using TF-IDF. In all algorithms either using TF-IDF or count vectorizer method set 4 have given better accuracy but poor performance on other three-parameter. Set 1 and set 2 given decent and stable values. But set 2 was better in all algorithms either using TF-IDF or the count vectorizer method. So, from our observation, we can say that an excessive amount of one type of data can be misleading while detecting fake news. For our further analysis, we have taken set 2, as our final comparison shown in Fig. 10. Using the TF-IDF method Multinomial Naïve Bayes does not show a satisfying result. But other three algorithms showed better outcomes Fig. 10b. On the other hand, using the count vectorizer method all of the four algorithms shown better outcomes as we can see in Fig. 10a

VI. CONCLUSION

Fake news is one of the alarming issues of the digital era and fake news detection can halt this issue. Through our work, we want to contribute to solve this issue by

Table 3: Fake News Detection Result Using Count Vectorizer

Technique	Algorithms		T/F	Support	Precision	Recall	F1-Score	Accuracy
Count Vectorizer	Multinomial naïve bias	Set-1	0	53	0.89	0.77	0.83	0.85
			1	63	0.83	0.92	0.87	
		Set-2	0	63	0.89	0.86	0.87	0.88
			1	69	0.89	0.86	0.87	
		Set-3	0	61	0.88	0.80	0.84	0.88
			1	104	0.89	0.93	0.91	
		Set-4	0	53	0.79	0.77	0.78	0.91
			1	211	0.94	0.95	0.95	
	Passive Aggressive	Set-1	0	53	0.89	0.89	0.89	0.90
			1	63	0.90	0.90	0.90	
		Set-2	0	63	0.89	0.86	0.87	0.88
			1	69	0.87	0.90	0.89	
		Set-3	0	61	0.84	0.84	0.84	0.88
			1	104	0.90	0.90	0.90	
		Set-4	0	53	0.80	0.68	0.73	0.9
			1	211	0.92	0.96	0.94	
	Support Vector Machine	Set-1	0	53	0.82	0.85	0.83	0.84
			1	63	0.87	0.84	0.85	
		Set-2	0	63	0.86	0.87	0.87	0.87
			1	69	0.88	0.87	0.88	
		Set-3	0	61	0.84	0.84	0.84	0.88
			1	104	0.90	0.90	0.90	
		Set-4	0	53	0.80	0.68	0.73	0.90
			1	211	0.92	0.96	0.94	
	Logistic Regression	Set-1	0	53	0.89	0.79	0.84	0.86
			1	63	0.84	0.92	0.88	
		Set-2	0	63	0.90	0.86	0.88	0.89
			1	69	0.88	0.91	0.89	
Set-3		0	61	0.88	0.82	0.85	0.89	
		1	104	0.90	0.93	0.92		
Set-4		0	53	0.92	0.62	0.74	0.91	
		1	211	0.91	0.99	0.95		

presenting a data set containing more than 2900 fake and true news on Covid-19. In this digital era, people are mostly dependent on online news sources than printed news. As there are lots of news sources available online. So, there is a high risk of getting fake news on the online platform and fake news detection using the machine learning techniques can reduce the immense spread. To boost solving this issue, we have collected a new data set from traditional media based on Covid-19. We hope that This data set will help future researchers to contribute more to detecting fake news. We have also analyzed our data by four different existing supervised algorithms. We have calculated accuracy, recall, f1-score, and precision regarding count vectorizer and TF-IDF to detect fake news using algorithms and our data set on Covid-19. We have analyzed the algorithm with the four different amount of data. We have also analyzed how excessive amounts of one type of data can severely affect detection systems. As in this work, we only focused on the news related to Covid-19 and the news are from the traditional media so for this the number of data decreased. A largescale data set can be more effective in building a fake news detection system. As the fake news collection is challenging itself, especially from traditional media as there is a little amount of fake

news in fact-checking sites. Also, we have faced similar kinds of fake news on different fact-checking sites. In the year 2020, many more events happened besides Covid-19. So, a positive path is to build a robust and large-scale data set of fake news from different media like social media, which researchers will use to facilitate additional study in this field. Using unsupervised or semi-supervised learning while detecting fake news can show better performance.

REFERENCES RÉFÉRENCES REFERENCIAS

1. Abdullah-All-Tanvir, Ehasas Mia Mahir, Saima Akhter, and Mohammad Rezwanul Huq. Detecting fake news using machine learning and deep learning algorithms. In 2019 7th International Conference on Smart Computing Communications (ICSCC), pages 1–5, 2019.
2. Vasu Agarwal, H. Parveen Sultana, Srijan Malhotra, and Amitrajit Sarkar. Analysis of classifiers for fake news detection. *Procedia Computer Science*, 165:377 – 383, 2019.
3. Sarah A. Alkhodair, Steven H.H. Ding, Benjamin C.M. Fung, and Junqiang Liu. Detecting breaking news rumors of emerging topics in social media.

- Information Processing and Management, 57(2):102018, 2020.
4. Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–236, May 2017.
 5. Amjad Atallah. Iran: Over 700 dead after drinking alcohol to cure coronavirus, 2020. <https://www.aljazeera.com/news/2020/4/27/iran-over-700-dead-after-drinking-alcohol-to-cure-coronavirus> accessed August 26, 2020.
 6. Pritika Bahad, Preeti Saxena, and Raj Kamal. Fake news detection using bi-directional LSTM-recurrent neural network. *Procedia Computer Science*, 165:74–82, 2019. 2nd International Conference on Recent Trends in Advanced Computing ICRTAC - DISRUP - TIV INNOVATION, 2019 November 11-12, 2019.
 7. Meital Balmas. When fake news becomes real: Combined exposure to multiple news sources and political attitudes of inefficacy, alienation and cynicism. *Communication Research*, 41, 07 2012.
 8. Paul Brewer, Dannagal Young, and Michelle Morreale. The impact of real news about "fake news": Intertextual processes and political satire. *International Journal of Public Opinion Research*, 25:323–343, 09 2013.
 9. Jason Brownlee. What is a Confusion Matrix in Machine Learning, 2020. <https://machinelearningmastery.com/confusion-matrix-machine-learning/> accessed August 22, 2020.
 10. Alistair Coleman. 'Hundreds dead' because of Covid-19 misinformation, 2020. <https://www.bbc.com/news/world-53755067> accessed August 22, 2020.
 11. Nadia K. Conroy, Victoria L. Rubin, and Yimin Chen. Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology*, 52(1):1–4, 2016.
 12. Corinna Cortes and Vladimir VAPNIK. Support vector networks. *Machine Learning*, 20:273–297, 09 1995.
 13. Shevon Desai. "Fake News," Lies and Propaganda: How to Sort Fact from Fiction, 2020. <https://guides.lib.umich.edu/c.php?g=637508&p=4462356> accessed December 05, 2020.
 14. Oberiri Destiny Apuke and Bahiyah Omar. Fake news and covid-19: modelling the predictors of fake news sharing among social media users. *Telematics and Informatics*, 2020.
 15. Alan Duke. Lead Stories, 2020. <https://leadstories.com/> accessed June 26, 2020.
 16. Ahmet Dursun. Iran: Death toll from toxic alcohol rises to 180, 2020. <https://www.aa.com.tr/en/health/iran-death-toll-from-toxic-alcohol-rises-to-180/1771659> accessed August 26, 2020.
 17. Mykhailo Granik and Volodymyr Mesyura. Fake news detection using naive bayes classifier. 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), pages 900–903, 2017.
 18. Mikell P. Groover. automation. *Encyclopedia Britannica.*, 2020. <https://www.britannica.com/technology/automation> accessed December 28, 2020.
 19. Nolan Higdon. The anatomy of fake news : a critical news literacy education. University of California Press, 2020. Accessed November 28, 2020.
 20. Angie Drobnic Holan. Politifact, 2020. <https://www.politifact.com/> accessed June 30, 2020.
 21. Junaed Younus Khan, Md. Tawkat Islam Khondaker, Sadia Afroz, Gias Uddin, and Anindya Iqbal. A benchmark study of machine learning models for online fake news detection. *Machine Learning with Applications*, 4:100032, 2021.
 22. Lauren Lutzke, Caitlin Drummond, Paul Slovic, and Joseph Arvai. Priming critical thinking: Simple interventions limit the influence of fake news about climate change on facebook. *Global Environmental Change*, 58:101964, 09 2019.
 23. Andrew McCallum and Kamal Nigam. A comparison of event models for naive bayes text classification. *Work Learn Text Categ.*, 752, 05 2001.
 24. Claire Milne. Full Fact, 2020. <https://fullfact.org/> accessed August 10, 2020.
 25. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
 26. Martin Porter. An algorithm for suffix stripping. *Program: electronic library and information systems*, 40:211–218, 07 2006.
 27. Anand Rajaraman and Jeffrey David Ullman. *Data Mining*. Cambridge University Press, 2011.
 28. Lori Robertson. Fact Check, 2020. <https://www.factcheck.org/>(accessed June 30, 2020).
 29. Natali Ruchansky, Sungyong Seo, and Yan Liu. CSI: A Hybrid Deep Model for Fake News Detection, page 797–806. Association for Computing Machinery, New York, NY, USA, 2017.
 30. Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall Press, USA, 3rd edition, 2009.
 31. Kanish Shah, Henil Patel, Devanshi Sanghvi, and Manan Shah. A comparative analysis of logistic regression, random forest and knn models for the text classification. *Augmented Human Research*, 5, 12 2020.
 32. Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake news detection on social

- media: A data mining perspective. SIGKDD Explor. Newsl., 19(1):22–36, September 2017.
33. Maciej Szpakowski. FakeNewsCorpus, 2020. <https://github.com/several27/FakeNewsCorpus> accessed November 23, 2020.
 34. Mark Trudeau. COVID-19 IMPACT: Broadband usage surges 47% IN Q1, nearing YE2020 expectations, with exponential rise in >1TB/2TB power users, 2020. <https://openvault.com/complimentary-report-Q12/accessed> August 22, 2020.
 35. Emmanuel Vincent. Health Feedback, 2020. <https://healthfeedback.org/> accessed July 04, 2020.
 36. William Yang Wang. “liar, liar pants on fire”: A new benchmark dataset for fake news detection”. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, page 422–426, 2017.
 37. Fan Yang, Arjun Mukherjee, and Eduard Dragut. Satirical news detection and analysis using attention mechanism and linguistic features. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 1979–1989. Association for Computational Linguistics, sep 2017.
 38. Chaowei Zhang, Ashish Gupta, Christian Kauten, Amit V. Deokar, and Xiao Qin. Detecting fake news for reducing misinformation risks using analytics approaches. European Journal of Operational Research, 279(3):1036–1052, 2019.
 39. Xichen Zhang and Ali A. Ghorbani. An overview of online fake news: Characterization, detection, and discussion. Information Processing and Management, 57(2):102025, 2020.



GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY: C
SOFTWARE & DATA ENGINEERING
Volume 22 Issue 2 Version 1.0 Year 2022
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Capillary X: A Software Design Pattern for Analyzing Medical Images in Real-time using Deep Learning

By Maged Abdalla Helmy Abdou, Paulo Ferreira, Eric Jul
& Tuyen Trung Truong

University of Oslo

Abstract- Recent advances in digital imaging, e.g., increased number of pixels captured, have meant that the volume of data to be processed and analyzed from these images has also increased. Deep learning algorithms are state-of-the-art for analyzing such images, given their high accuracy when trained with a large data volume of data. Nevertheless, such analysis requires considerable computational power, making such algorithms time- and resource-demanding. Such high demands can be met by using third-party cloud service providers. However, analyzing medical images using such services raises several legal and privacy challenges and do not necessarily provide real-time results. This paper provides a computing architecture that locally and in parallel can analyze medical images in real-time using deep learning thus avoiding the legal and privacy challenges stemming from uploading data to a third-party cloud provider. To make local image processing efficient on modern multi-core processors, we utilize parallel execution to offset the resourceintensive demands of deep neural networks. We focus on a specific medical-industrial case study, namely the quantifying of blood vessels in microcirculation images for which we have developed a working system.

GJCST-C Classification: DDC Code: 020.3 LCC Code: Z1006



Strictly as per the compliance and regulations of:



© 2022. Maged Abdalla Helmy Abdou, Paulo Ferreira, Eric Jul & Tuyen Trung Truong. This research/review article is distributed under the terms of the Attribution-NonCommercial-NoDerivatives 4.0 International (CC BYNCND 4.0). You must give appropriate credit to authors and reference this article if parts of the article are reproduced in any manner. Applicable licensing terms are at <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

Capillary X: A Software Design Pattern for Analyzing Medical Images in Real-time using Deep Learning

Maged Abdalla Helmy Abdou ^α, Paulo Ferreira ^σ, Eric Jul ^ρ & Tuyen Trung Truong ^ω

Abstract- Recent advances in digital imaging, e.g., increased number of pixels captured, have meant that the volume of data to be processed and analyzed from these images has also increased. Deep learning algorithms are state-of-the-art for analyzing such images, given their high accuracy when trained with a large data volume of data. Nevertheless, such analysis requires considerable computational power, making such algorithms time- and resource-demanding. Such high demands can be met by using third-party cloud service providers. However, analyzing medical images using such services raises several legal and privacy challenges and do not necessarily provide real-time results. This paper provides a computing architecture that locally and in parallel can analyze medical images in real-time using deep learning thus avoiding the legal and privacy challenges stemming from uploading data to a third-party cloud provider. To make local image processing efficient on modern multi-core processors, we utilize parallel execution to offset the resourceintensive demands of deep neural networks. We focus on a specific medical-industrial case study, namely the quantifying of blood vessels in microcirculation images for which we have developed a working system. It is currently used in an industrial, clinical research setting as part of an e-health application. Our results show that our system is approximately 78% faster than its serial system counterpart and 12% faster than a master-slave parallel system architecture.

1. INTRODUCTION

Early attempts to address the problem of running demanding computational algorithms in tightly constrained environments emerged in the 1980s [1]. Performance limitations became apparent with the rise of processing big data using deep learning (DL) techniques because DL requires large amounts of computational power [2], [3]. Such limitations included the under-utilization of the available computing resources to execute processes introducing undesirable delays [4]. These limitations are still prominent when real-time results are desired in tightly constrained environments (i.e., clinical environments). Furthermore, using third-party cloud services to rent computing resources is risky due to General Data Protection Regulation (GDPR) [5]. These regulations effectively limit clinicians to local computing resources, such as laptops and PCs approved for use at hospitals.

Author ^α ^σ ^ρ: Department of Informatics University of Oslo, Norway.
e-mail: magedaa@uio.no

Author ^ω: Department of Mathematics University of Oslo, Norway.

This paper aims to design, implement, and evaluate a software package that can analyze medical images using deep learning in a local environment as to mitigate the risk of breaching GDPR rules while still getting results in real-time. We focus on a specific industrial, medical case study: the quantification of blood vessels in microcirculation images captured by using in-clinic, hand-held cameras with microscope lenses. The quantified value is called capillary density or blood vessel density. This value is of high clinical relevance because the fluctuation of this value can be used as an early marker to indicate an organ failure, and the severity of the change might predict the chances of the patient surviving [6]–[12].

The requirements of our system were established by interviewing a set of medical doctors and surgeons who spent several years in the microcirculation analysis field (associated with ODI Medical AS, a MedTech company responsible for the e-health industrial application). The main requirements for a production-grade system for the quantification of blood vessels analysis captured by a real-time camera are:

1. The system must be able to analyze a microcirculation image (1920x1080) in real-time (one second or less);
2. The system must have low power consumption so that it can be used in battery-powered devices in hospitals; and
3. The system must be built on top of a popular, widely used programming language and framework (e.g., Python and Tensorflow) running on standard hardware.

To the best of our knowledge, no previous work on microcirculation analysis reported using parallel frameworks to calculate the capillary density in under 1 second for a frame with a resolution of 1920x1080 on a CPU using deep learning with an accuracy of ~85%. The medical doctors proposed this accuracy value to outperform the accuracy achievable by a trained clinician. Previous systems in the literature that achieve a comparable or higher accuracy needed a GPU that is not available in typical low-power computers approved for use in hospitals. The developed system runs in an industrial, clinical environment on a standard low cost

computer utilizing all the available resources and meets the requirements listed above. This paper does not focus on developing the deep learning algorithm but rather the deployment of the deep learning algorithm. The algorithm used in this paper achieves an accuracy of ~85%, and is described in a previous paper [13].

In Section II, we present the work related to our paper including a literature survey on the relevant parallel frameworks and existing systems that were built to analyze micro-1 circulation images. In Section III, we present the proposed architecture for our package along with two baseline systems that we used to benchmark our proposed architecture against. In Section IV, we present how we implemented our system. In Section V, we present the evaluation criteria that have been used to evaluate our system and benchmark our proposed system against a baseline serial system and a baseline parallel system with the presented criteria and discuss our results. In Section VI, we present our conclusion.

II. RELATED WORK

This section presents the literature review on current parallel frameworks and existing systems built to calculate capillary density.

a) Parallel Frameworks

Hadoop [14] gained recognition in 2004 and provides a framework for distributed storage and the processing of big data. It splits large blocks of data into a Hadoop Distributed File System (HDFS) which is based on Google's file system (GFS) and stores data across clusters [15], [16]. HDFS uses data locality, allowing clusters and nodes to manipulate data, making it faster than conventional high-performance computing [17].

MapReduce then processes the data stored on HDFS [18]. MapReduce has a master job tracker and one per cluster to schedule jobs, manage resources, and re-execute processes when a node fails [17]. HDFS and MapReduce are two modules built to store and process big data reliably. However, the main drawback of Hadoop is that it cannot deal with big data real-time stream processing; therefore, Apache Spark was introduced [19] was introduced in 2010.

Some benchmarks show that Spark is three times faster than Hadoop [20]. This increase is because Spark can load and process data using RAM instead of the two-stage access paradigm introduced by MapReduce [19]. Spark outshines MapReduce when it comes to real-time processing [21], [22]. Furthermore, the ease of programming on Spark with Scala [23], Java [24] and Python [22] makes it relatively easy to adapt instead of MapReduce, which can be programmed only in Java. Spark provides a unified processing system instead of several isolated applications that do not share the state amongst each other [25]. Although Spark was

designed to outperform MapReduce processing, its the fundamental limitation is the complexity involving asynchronous execution and the compatibility issues introduced when integrating it into the deep learning lifecycle [26].

Dask [27] was introduced in 2014 and is a parallel computing library that uses dynamic task scheduling to leverage multi-core processors and High-Performance Computing (HPC) clusters. Instead of loading all data into RAM, Dask pulls data into RAM in chunks and throws away intermediate values as soon as possible, freeing more memory to process more data [27]. While Spark can be seen as an extension to the MapReduce paradigm, Dask is a generic task scheduling system that handles complex dimensional arrays [28]–[30]. Both Dask and Spark leverage acyclic graphs, but the map stage of Dask can represent more complex algorithms than Spark [31]. Thus, Dask can parallelize sophisticated algorithms without excess memory usage [29]. Moreover, Spark does not natively support multi-dimensional arrays as Dask does [30], [32]. This advantage makes Dask lightweight and smaller than Spark, and Dask integrates natively with the numeric Python ecosystem. However, Dask is not fully compatible with TensorFlow, and deep learning algorithms as the framework focuses on Data science libraries like Pandas and Numpy.

Orleans [33] is an actor system that provides highly available concurrent distributed systems. The main drawback is how the system reacts to a data failure event. Developers must manually create checkpoint actor states and intermediate responses to restore stateful actors [34]. While this does not affect the performance, it can bring some overhead when developing a system to handle failure events.

Tensorflow [35] is an ecosystem of machine learning and deep learning tools that leverage CPUs and GPUs while training. However, it provides limited support when deploying it to serve users because it does not fully support responses when a task is completed or when a fault is detected. One way to perform this activity is to wrap the Tensorflow Model in a flask service and serve the model [36]. However, this becomes unmanageable when scaling with different models. Tensorflow serving [37] was introduced to deploy models in production environments but has to be used in conjunction with traditional web servers, which introduces additional latency.

With the introduction of deep learning techniques [38], which consists of several millions of parameters to compute, wrapping deep learning models with traditional servers is no longer sufficient. Compared to traditional models, deep learning models are computationally intensive and have a response time of tens of a millisecond or greater [39]; thus, there is a need for efficient parallelizing to reduce the response time.

Frameworks such as MapReduce [17] and Spark [40] are not suitable for models serving in real-time because they were designed and built for batch processing. Furthermore, they are not suitable for large numbers of small transactions because of the considerable time overhead that they require for instantiation. Dask [27] and Tensorflow [35] provide a complex and very little support for model serving [26].

It is possible to set up different parts of different frameworks together to have a system that can serve a deep learning model. However, the compatibility and maintenance of these different frameworks increase the technical complexity. Unfortunately, deploying deep

learning models into production is still not a straightforward endeavor.

b) Existing Microcirculation Analysis Systems

This section presents the current work on systems that calculate capillary density from microcirculation images.

As briefly mentioned at the end of the introduction, none of the existing works mentioned on microcirculation analysis reported using parallel frameworks to calculate the capillary

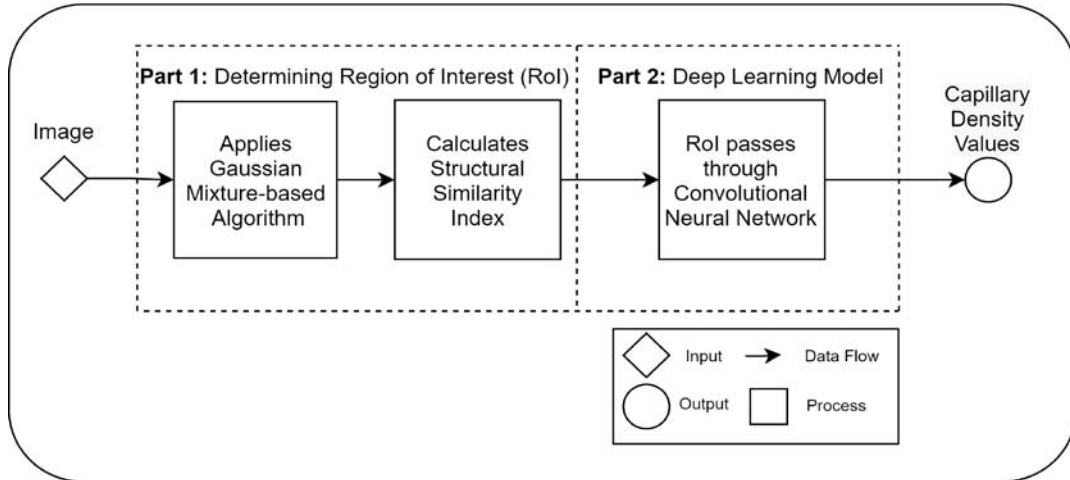


Fig. 1: The diagram shows the code encapsulated in a core on a computer. This code is replicated across each core to achieve parallelism. This code calculates the capillary density from a microcirculation image. The architecture consists of two parts: i) first determining the RoI using traditional computer vision algorithms and ii) then using deep learning to classify if the RoI contains a capillary

density in under ~1 second for a frame with a resolution of 1920x1080 on a CPU using deep learning with an accuracy of ~85%. Those who exceeded this accuracy used a GPU which is not readily available in a clinical environment.

Cynthia Cheng et al. [41] takes a three-step approach to quantify capillary density. First, they apply an image enhancement process to darken the capillaries and lighten the background. They then flatten the image using 2D filters and raise the image's contrast. The image is then despeckled using a 7x7 filter. They then adjust the histogram of the image to a best-fit model. The second step involves manually selecting the capillary as a target object. They then select the background as a reference. The algorithm then selects the rest of the capillaries and excludes the images. A macro is then created from this process, which can be applied to other images with similar characteristics. As described, this involves several steps, including the manual user intervention; therefore cannot provide results in less than 1 second.

A. Tama et al. [42] uses binarization followed by skeleton extraction and segmentation to quantify the capillaries. The first step involves extracting a reference

image. The image has to be then manually cropped by the user. The green channel is then extracted from the image to have the highest probability of vessels in it. They then apply a top-hat transform to remove unevenness in the background. They then apply the Wiener filtering, a lowpass filter followed by Gaussian smoothing. They then apply Otsu thresholding to segment the image from the background and apply a skeleton extraction method to quantify the capillary. The authors do not report the speed needed to perform these steps.

Sherry G. Clendenon et al. [43] uses a manual method to segment the microvascular structure. The authors do not report the speed or accuracy of their method.

Pavle Prentić et al. [44] used a custom neural network to segment the foveal microvasculature. Their neural network consists of three Convolutional Neural Network (CNN) blocks coupled with max-pooling and a dropout layer followed by two dense layers. They reported accuracy of 82.4% at 2 minutes.

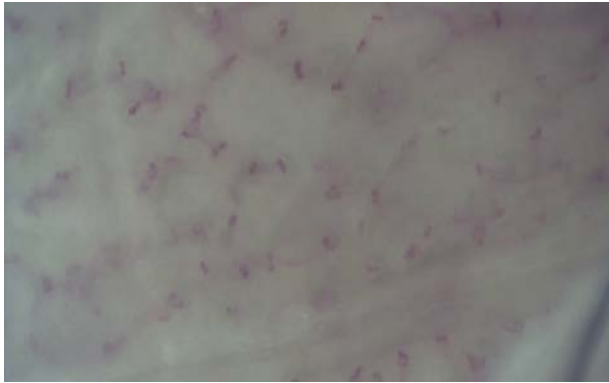
R Nivedha et al. [45] used a non-linear Support Vector Machine [46] to classify images. They first started by extracting the green channel since it contains the

relevant information to detect blood vessels. They then performed manual cropping and used adaptive histogram equalization to improve the image's contrast. They then used image enhancement to segment the image using a Gaussian filter followed by OTSU thresholding. They then used PrincipalComponent Analysis(PCA) to extract the features. A Support Vector Machine then performed the classification. They reported accuracy of 83.3% but not the time needed for automated analysis.

KV Suma et al. [47] used Fuzzy Logic Kernels to classify the images. They started by Fuzzification of the input, followed by the Application of the Fuzzy operator, then aggregating the consequents across the rules,

ending with results. They reported an overall accuracy of 83.3% but not the time needed for automated analysis. In their next paper [48], they experimented with different types of machine learning techniques, including Random Forests Classifier, Multinomial Logistic Regression, and CNNs. However, they do not report the timing needed for classifying the blood vessels.

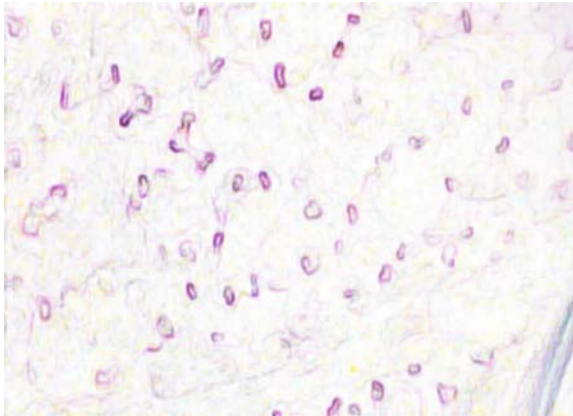
Perikumar Java et al. [49] used a custom form of ResNet18 [50] to quantify capillaries. They used a 10-layer architecture and resized the images to input 224x224x3. They



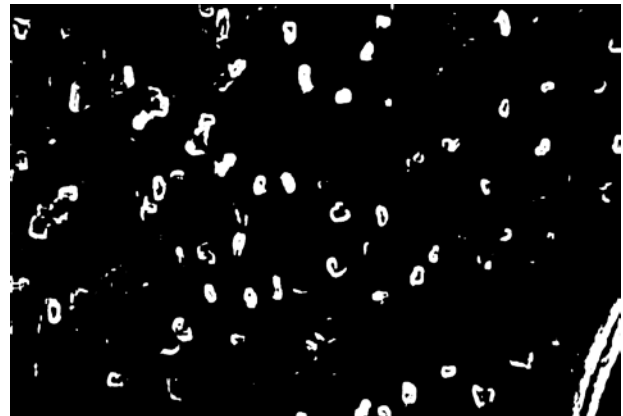
(a)



(b)



(c)



(d)

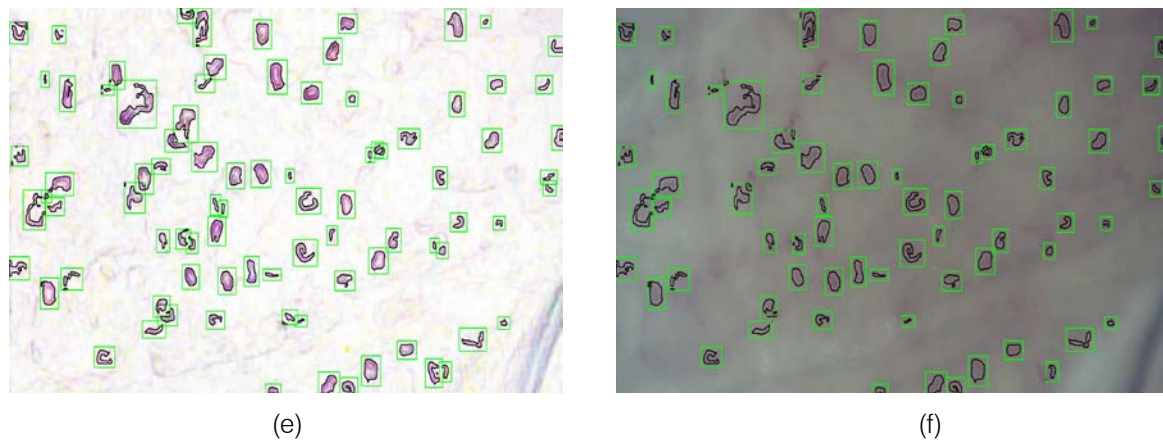


Fig. 2: (a) This presents a sample of a microcirculation image that is taken as an input to the system (b) The background image calculated using a Gaussian Segmentation Algorithm (c) The segmented area formed by calculating the difference between the original image and the background (d) The Structural Similarity Index calculated from the original frame and the background image (e) The modified image with the capillary area highlighted in black encapsulated within the green bounding box (f) The original image with the capillary area highlighted in black encapsulated within the a bounding box

applied the Adam optimizer and cross-entropy as loss metrics. They trained the NVIDIA GeForce GTX Titan X algorithm and used the PyTorch library. They reported accuracy of 88%. However, such an algorithm is not suitable for a clinical environment due to the high-end GPU required to run it.

F Ye et al. [51] utilized the concept of transfer learning and used the Inception Single Shot Multibox Detector (SSD) [52] to build their neural network. They build their system using Python and Tensorflow with an image resolution of 744×482 pixels. They applied data augmentation to the image to increase the number of datasets. The SSD architecture requires GPU to produce results in real-time, making it unsuitable to be used in a clinical requirement with only CPUs available.

YS Hariyani et al. [53] used U-net architecture combined with a dual attention module. They introduced a new method called DA-CapNet, which can analyze microcirculation images. It consists of the encoder and decoder parts. The encoder downsamples the dimension of the information in an image while increasing the number of channels. This step increases the spatial information dimension. They then combine it with a dual attention module which increases the accuracy. The dual attention uses the squeeze and excitation process to extract the blood vessels in the image. The authors resized the image to 256×256 to reduce the processing time and used a Gaussian threshold method with a median blurring filter of kernel size five. The authors reported accuracy of 64% but not the time taken for analyses.

G Dai et al. [54] used a custom neural network similar to Pavle Prentić et al. for segmentation. However, G Dai et al. used five CNN blocks instead of three. Hang-Chan Jo et al. [55] used a Attention-UNet architecture [56]. Their method starts by using the

CLAHE method and computes several histograms. They then apply the Gamma correction and pass it to the deep neural network. The reported accuracy was 73.20%, but not the time is taken for analysis.

III. PROPOSED SYSTEM

This section presents the system's architecture to analyze medical images in parallel, specifically, to calculate the capillary density in a microcirculation image. We start by presenting the DL part (which is based on OpenCV [57] and Tensorflow [35]) and the architecture of our system's parallel part (which is based on Ray [58]).

a) *The Deep Learning Algorithm part of the Proposed System*

The outline of the deep learning architecture is shown in Figure 1. It consists of two main parts: i) determining the regions of interest (RoIs) where capillaries might exist, and ii) using a CNN for predicting whether these RoIs contain a capillary or not.

The original frame is shown in Figure 2a. The position of the capillaries is determined by first removing the background from the original frame using a Gaussian Mixture-based Background/Foreground Segmentation Algorithm [59]. The background removed is shown in Figure 2b. The structural similarity index measure (SSIM) [60], [61] is applied between the original frame shown in Figure 2a and background image shown in Figure 2b resulting in Figure 2c and Figure 2d. Bounding boxes are formed around the red areas using OpenCV contour method [62]. These bounding boxes are then passed to the CNN for prediction. The RoIs that have been predicted as capillaries have a green bounding box around each one of them along with a black line to highlight the shape of

the capillary. This is shown in Figure 2e and the original image in Figure 2f. The number of pixels within the encapsulated black contour line is summed up and divided by the total number of pixels resulting in the value of the capillary density.

Bounding boxes are formed around the predicted bounding boxes using the OpenCV contour method [62]. These bounding boxes are then passed to the CNN for prediction. The Rols predicted as capillaries have a green bounding box around them and a black line to highlight the capillary shape. This is shown in Figure 2c and Figure 2d. The number of pixels within the encapsulated black contour line is summed

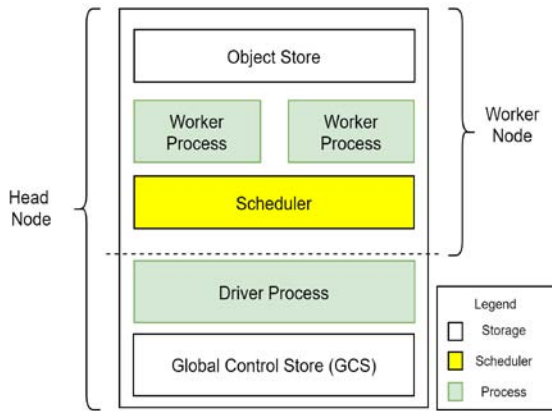


Fig. 3: A Breakdown of the Building Blocks Used to Built the Proposed System

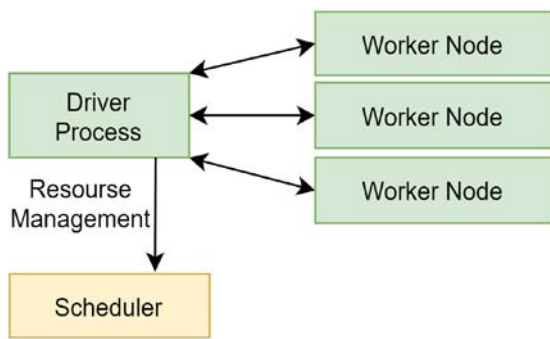


Fig. 4: The data flow view of how the driver process coordinates with the driver process and the workers

up and divided by the total number of pixels resulting in the value of the capillary density.

The CNN consisted of three blocks of Conv2D. The first Conv2D consisted of 32 filters, the second Conv2D consisted of 64 filters, and the third Conv2D consisted of 128 filters with a block of Maxpooling2D. All the Conv2D blocks have a filter of 3x3 shape. Two dense layers of 128 neurons follow the Conv2D blocks, 64 neurons, and two neurons. The Rectified Linear Unit (ReLu) [63] activation function is used for the whole network except the last neuron layer, which used a softmax activation function [64]. This network has been

trained on ~11,000 images of capillaries captured by trained professionals in a clinical setting¹. The details and the specificity of the algorithms and data used to train the algorithm can be found in a previous paper by the same authors [13].

b) *The Parallel System part of the Proposed System*

This architecture has two types of nodes: the worker nodes and a head node. A worker node consists of the worker process(es), the scheduler, and the object-store. A worker node and a head node anatomy are shown in Figure 3 and the data flow within the components is shown in Figure 4.

A worker process encapsulates the code to be executed and is responsible for task submission and execution of tasks. In our system, the worker node encapsulates the deep learning algorithms. It receives the image to be analyzed and replies whether this image contains a capillary (blood vessel) or not. The scheduler is the resource manager of the worker node. The object store stores and transfers object larger than 100KB. The head node has a Global Control Store (GCS) and a driver process. The GCS is a key-value server that contains objects, actors, and tasks. The driver process submits tasks to the scheduler and keeps track of the objects created with all the nodes. When the code is initiated, an instance of a head node is created. The maximum number of worker processes within this head node is based on the number of parallel modules in the architecture instantiated and the maximum number of cores. Each worker performs both stages: suggesting Rols and detecting capillaries using the CNN loaded. Each worker returns a single object that contains the frame's density value and is stored in the object-store. The code execution of this architecture is scheduled using the scheduler, and the tasks are performed over a general-purpose Remote Procedure call to the worker processes on top of the Python interpreter. The scheduler then communicates the results via an object transfer protocol. For error handling and fault tolerance, the scheduler retries executing it on the worker processor, if a task fails due to a worker process ending unexpectedly.

Thus, one of the main differences between the proposed system and the baseline parallel system is that the former uses a driver process to manage the workers while the latter uses a controller and a router to manage the worker's tasks. A baseline parallel system uses some controller and router to prevent the worker's potential overloading with tasks, which can cause it to fail. However, these two components (controller and router) can occupy up to two cores for the management of the workers without performing any code execution.

¹ The sponsoring company provided a device that was used to capture this data.

While the proposed system does not reserve any cores to manage the drivers but rather re-executes the code if a worker fails [65].

When the code is instantiated in our proposed system, the worker node loads the CNN as a Tensorflow model. Each worker occupies a logical processor, thread, or core, depending on the CPU architecture; we assume it is a core and instantiate a worker per core. As the number of cores increased, the number of images processed in parallel increased with the number of cores.

We have shown that by combining the deep neural network part with the parallel part, we can process several images at the same time, suggest RoIs and predict whether the bounding boxes have a capillary or not. Furthermore, the number of frames processed in parallel is determined by the maximum number of cores available or the pre-defined the value inserted by the user (assuming it does exceed the number of cores available).

IV. IMPLEMENTATION

Many programming languages can implement a parallel processing framework. Python is the fastest-growing programming language [66], [67] and the preferred programming language for deep learning with Tensorflow [68], [69]. This popularity stems from its design philosophy, where it emphasizes readability and simplicity [66]. Moreover, the number of libraries, various tools, and speedily expanding the industrial community supporting Python made the language attractive [70].

Thus, the proposed package was built on top of Python 3.7 [22], OpenCV 4.5.2[57], Scikit-learn 0.18[71], Ray 1.2[26] and Tensorflow 2.3[35]. The coding and evaluation were done in Pycharm Professional 2021.1 on a Windows 10 operating system. The system can be installed, modified, and used by following the instructions in the readme file on the Github repository (www.github.com/magedhelmy1/CCGRID 2022 parallel system for image analysis).

To use the system, the user can clone the package from the Github repository and import it in their Python environment.

V. EVALUATION AND DISCUSSION

In this section, we compare the baseline serial architecture, the baseline parallel architecture, and the proposed system with each other using the following three metrics: execution time, speedup, and CPU usage. We show that the proposed Python system is 78% faster than the serial system and 12% faster than the baseline parallel architecture. These three metrics are standardized markers to quantify a system performance [72]. We use these three metrics to compare our proposed approach to a serial and parallel

system with the same deep neural network. We show that the proposed system meets the requirements mentioned in Section I and supersedes both the baseline serial system and the baseline parallel system in execution time, speedup, and CPU usage. The proposed system, serial counterpart, and parallel counterpart had the same CNN model and microcirculation images. We evaluated the three systems by taking the average time to calculate capillary density per image for a set of 100 images, which is an arbitrary number we chose to reduce the margin of error and ensure our calculations' accuracy.

a) Execution Time

To calculate how much one architecture was faster compared to the other, we used Equation 1, where ET denotes execution time.

$$\frac{SlowerET - FasterET}{SlowerET} = \%Faster \quad (1)$$

The execution time metric measures the average time needed to calculate a single image's capillary density. We used 100 images in each architecture to reduce the measurement error margin. The execution time of each architecture is the following:

1. Baseline serial architecture — one second per frame;
2. Baseline parallel architecture — 0.25s per frame; and
3. The proposed system — 0.22s per frame. The average

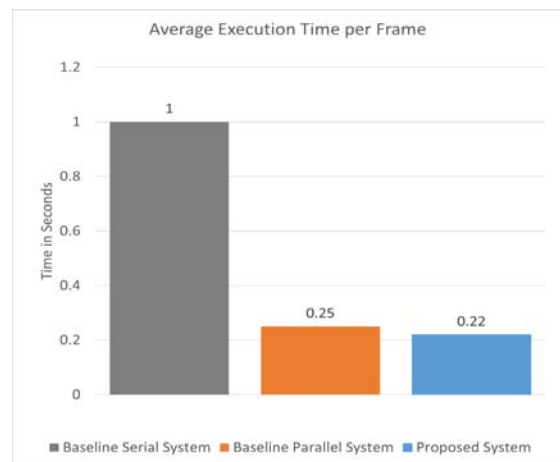


Fig. 5: The execution time of the proposed system against the baseline serial system and the baseline parallel system

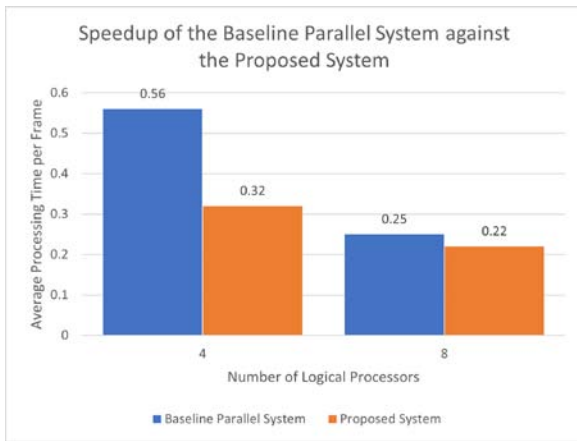


Fig. 6: This graph shows the speedup of the baseline parallel system and the proposed system as the number of cores increases

values were calculated by measuring the time to process a frame in a set of 100 microcirculation images. The execution time of the three architectures is presented in Figure 5.

In short, our results show that our proposed system is 12% faster than the baseline parallel architecture and 78% faster than its baseline serial architecture.

b) Speedup

This metric calculates the speed gain by the system as the number of cores increases. For the baseline serial architecture, the execution time is one second regardless of the number of cores available (indicating that the system is not scalable). The average execution time of processing one frame for the baseline parallel system and the proposed system is shown in Figure 6.

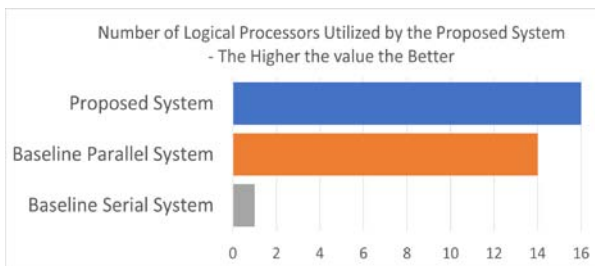


Fig. 7: This graph shows the number of cores used by each architecture to process a frame. The more used at any instance the better since this shows how efficient the system is at utilizing all the resources available to it.

The baseline parallel system processed a frame on average in 0.56 seconds with four cores, while the proposed system processed a frame in 0.32 seconds. The proposed system processes a frame 68% faster than the baseline serial architecture and ~43% faster than the baseline parallel architecture. As the number of cores doubles, the proposed system gains an additional ~31% during the baseline parallel architecture gains an

additional ~55%. In both cases, the proposed system outperforms the baseline parallel architecture. One of the main reasons the proposed system outperformed a baseline parallel system with a master-slave architecture is that a masterslave architecture can reserve up to two cores to manage the other parallel cores. In contrast, the proposed system does not reserve any cores beforehand. In this way, we free up the computer cores to focus on processing images rather than purely handling requests. Thus, the proposed system gains more speedup than the baseline parallel system. We can conclude that the proposed system has the recommended architecture for running deep neural networks on a single machine.

c) System Resource Utilization - CPU Usage

This metric measures the number of cores used to process the medical images using deep learning.

With the baseline serial architecture, only one core is utilized per frame due to the Python Global Interpreter Lock's limitation. With the baseline parallel architecture, it is always two less than the available number of cores because it always reserves these two for the management of the parallel workers. Each cores in the proposed system is allocated a task where each task processes a frame. Thus, the proposed system is most efficient on a single machine with a multi-core. A graph showing the number of cores used by each architecture is shown in Figure 7.

d) System Generalization

Our system functions and classes were built using modular design patterns. This design philosophy means that the user can replace the DL part of our system with their algorithm by simply pointing the function in our code to their algorithm. The details of this are highlighted in the README file in the GitHub repository. Thus, our package can be generalized to analyze images using a DL model of the user's choice in parallel. The system will automatically scale to the number of cores available without the user having to worry about experiencing issues with dependency, integration, resource utilization, and speedup.

VI. CONCLUSION

This paper presented a software package that can analyze medical images using DL locally. Our proposed system can efficiently use all local resources because it utilizes parallel execution to offset the resource-intensive demands of using a deep neural network. The proposed system is of high clinical relevance because monitoring changes in capillary density can be used to locate early markers indicating organ failure. The severity of the change in capillary density might predict whether or not the patient survives. Furthermore, clinical researchers do not risk uploading patient data to a third-party cloud provider to use a deep neural network to automatically analyze their images.

Our experiments show that our system provides an optimal design for using deep learning models running on a multicore single machine for image analysis. We benchmarked our system with a baseline serial architecture and a baseline parallel architecture using standardized evaluation metrics: execution time, speedup, and CPU usage. These metrics are used to calculate the performance of a system. Our results indicate that the proposed system is approximately 78% faster than its baseline serial system counterpart and 12% faster than a baseline parallel system.

As demonstrated by our evaluation criteria, our system exhibits an acceptable industrial performance compared to the other two presented baseline systems. This argument is further strengthened because our system is currently used as a product in an industrial setting to calculate and track capillary changes in patients with pancreatitis, COVID-19, and acute heart diseases. The clinical researchers welcomed using this system to analyze their medical images locally. This acceptance was mainly due to the system reducing analysis time and removing the risks of uploading the data to a thirdparty cloud provider.

Our code has been made public as an open-source project in a GitHub repository for testing and usage by other clinical users. The users can import the package into their Python environment and immediately start using it. Moreover, users who can clone the code from GitHub can swap our algorithm with theirs, showing that our architecture can be generalized and utilized in the context of other use cases that require image analysis running on a CPU in near real-time. Thus, the generality of our approach can be justified by several other use cases that require image analysis.

ACKNOWLEDGMENT

The authors would like to thank the Research Council of Norway for providing the necessary funds for this project. The research carried out was funded under these two projects; Industrial Ph.D. project nr: 305716 and BIA project nr: 282213. We would also like to thank ODI Medical AS for providing the requirements, testing the system, and integrating it as part of their e-health application.

REFERENCES RÉFÉRENCES REFERENCIAS

1. R. Duncan, "A survey of parallel computer architectures," *Computer*, vol. 23, no. 2, pp. 5–16, 1990.
2. Q. Zhang, L. T. Yang, Z. Chen, and P. Li, "A survey on deep learning for big data," *Information Fusion*, vol. 42, pp. 146–157, 2018.
3. N. C. Thompson, K. Greenewald, K. Lee, and G. F. Manso, "The computational limits of deep learning," *arXiv preprint arXiv:2007.05558*, 2020.
4. S. Pumma, M. Si, W.-c. Feng, and P. Balaji, "Parallel i/o optimizations for scalable deep learning," in *2017 IEEE 23rd International Conference on Parallel and Distributed Systems (ICPADS)*, IEEE, 2017, pp. 720–729.
5. J. M. M. Rumbold and B. Pierscionek, "The effect of the general data protection regulation on medical research," *Journal of medical Internet research*, vol. 19, no. 2, e47, 2017.
6. M. J.S. Parker and N. W. McGill, "The established and evolving role of nailfold capillaroscopy in Connective-Tissue disease," in *Connective Tissue Disease – Current State of the Art [Working Title]*, IntechOpen, 2019.
7. V. Nama, J. Onwude, I. T. Manyonda, and T. F. Antonios, "Is capillary rarefaction an independent risk marker for cardiovascular disease in south asians?" en, *J. Hum. Hypertens.*, vol. 25, no. 7, pp. 465–466, Jul. 2011.
8. B. Fagrell and M. Intaglietta, "Microcirculation: Its significance in clinical and molecular medicine," *Journal of internal medicine*, vol. 241, no. 5, pp. 349–362, 1997.
9. A. J. H. M. Houben, R. J. H. Martens, and C. D. A. Stehouwer, "Assessing microvascular function in humans from a chronic disease perspective," en, *J. Am. Soc. Nephrol.*, vol. 28, no. 12, pp. 3461–3472, Dec. 2017.
10. T. Wester, Z. A. Awan, T. S. Kvernebo, G. Salerud, and K. Kvernebo, "Skin microvascular morphology and hemodynamics during treatment with venoarterial extra-corporeal membrane oxygenation," en, *Clin. Hemorheol. Microcirc.*, vol. 56, no. 2, pp. 119–131, 2014.
11. A. L´opez et al., "Effects of early hemodynamic resuscitation on left ventricular performance and microcirculatory function during endotoxic shock," en, *Intensive Care Med Exp*, vol. 3, no. 1, p. 49, Dec. 2015.
12. G. Hernandez, A. Bruhn, and C. Ince, "Microcirculation in sepsis: New perspectives," *Current vascular pharmacology*, vol. 11, no. 2, pp. 161–169, 2013.
13. M. Helmy, A. Dykyy, T. T. Truong, P. Ferreira, and E. Jul, "Capillarynet: An automated system to analyze microcirculation videos from handheld vital microscopy," *arXiv preprint arXiv: 2104.11574*, 2021.
14. A. Holmes, *Hadoop in practice*. Manning New York; 2012, vol. 3.
15. D. Borthakur et al., "Hdfs architecture guide," *Hadoop Apache Project*, vol. 53, no. 1-13, p. 2, 2008.
16. K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The hadoop distributed file system," in *2010 IEEE 26th symposium on mass storage systems and technologies (MSST)*, IEEE, 2010, pp. 1–10.

17. J. Dean and S. Ghemawat, "Mapreduce: Simplified data processing on large clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, 2008.
18. —, "Mapreduce: A flexible data processing tool," *Communications of the ACM*, vol. 53, no. 1, pp. 72–77, 2010.
19. M. Saouabi and A. Ezzati, "A comparative between hadoop mapreduce and apache spark on hdfs," in *Proceedings of the 1st International Conference on Internet of Things and Machine Learning*, 2017, pp. 1–4.
20. S. Gopalani and R. Arora, "Comparing apache spark and map reduce with performance analysis using kmeans," *International journal of computer applications*, vol. 113, no. 1, 2015.
21. K. Aziz, D. Zaidouni, and M. Bellafkih, "Real-time data analysis using spark and hadoop," in *2018 4th international conference on optimization and applications (ICOA)*, IEEE, 2018, pp. 1–6.
22. G. Van Rossum et al., "Python programming language," in *USENIX annual technical conference*, vol. 41, 2007, p. 36.
23. M. Odersky et al., "An overview of the scala programming language," 2004.
24. K. Arnold, J. Gosling, and D. Holmes, *The Java programming language*. Addison Wesley Professional, 2005.
25. A. Spark, "Apache spark," Retrieved January, vol. 17, p. 2018, 2018.
26. P. Moritz et al., "Ray: A distributed framework for emerging ai applications," in *Symposium on Operating Systems Design and Implementation*, 2018, pp. 561–577.
27. M. Rocklin, "Dask: Parallel computation with blocked algorithms and task scheduling," in *Proceedings of the 14th python in science conference*, Citeseer, vol. 126, 2015.
28. M. Dugr'e, V. Hayot-Sasson, and T. Glatard, "A performance comparison of dask and apache spark for dataintensive neuroimaging pipelines," in *2019 IEEE/ACM Workflows in Support of Large-Scale Science (WORKS)*, IEEE, 2019, pp. 40–49.
29. P. Mehta et al., "Comparative evaluation of big-data systems on scientific image analytics workloads," *arXiv preprint arXiv:1612.02485*, 2016.
30. D. O. W. .-. D. S. analytics in Python. "Dask document-tation - comparison to spark." (Mar. 2021), [Online]. Available: <https://docs.dask.org/en/latest/spark.html>.
31. R. Nishihara et al., "Real-time machine learning: The missing pieces," in *Proceedings of the 16th Workshop on Hot Topics in Operating Systems*, 2017, pp. 106–110.
32. H. Li, G. Fox, and J. Qiu, "Performance model for parallel matrix multiplication with dryad: Dataflow graph runtime," in *2012 Second International Conference on Cloud and Green Computing*, IEEE, 2012, pp. 675–683.
33. S. Bykov, A. Geller, G. Kliot, J. R. Larus, R. Pandya, and J. Thelin, "Orleans: Cloud computing for everyone," in *Proceedings of the 2nd ACM Symposium on Cloud Computing*, 2011, pp. 1–14.
34. P. Bernstein, S. Bykov, A. Geller, G. Kliot, and J. Thelin, "Orleans: Distributed virtual actors for programmability and scalability," *MSR-TR-2014-41*, 2014.
35. T. O. Webpage. "Tensorflow." (Dec. 2021), [Online]. Available: <https://www.tensorflow.org/>.
36. F. A. Aslam, H. N. Mohammed, and P. Lokhande, "Efficient way of web development using python and flask," *International Journal of Advanced Research in Computer Science*, vol. 6, no. 2, pp. 54–57, 2015.
37. C. Olston et al., "Tensorflow-serving: Flexible, high-performance ml serving," *arXiv preprint arXiv:1712.06139*, 2017.
38. G. Litjens et al., "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.
39. S. Han, "Efficient methods and hardware for deep learning," Ph.D. dissertation, Stanford University, 2017.
40. M. Zaharia et al., "Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing," in *9th Symposium on Networked Systems Design and Implementation 12)*, 2012, pp. 15–28.
41. C. Cheng, C. W. Lee, and C. Daskalakis, "A reproducible computerized method for quantitation of capillary density using nailfold capillaroscopy," *Journal of visualized experiments: JoVE*, no. 104, 2015.
42. A. Tama, T. R. Mengko, and H. Zakaria, "Nailfold capillaroscopy image processing for morphological parameters measurement," in *2015 4th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME)*, IEEE, 2015, pp. 175–179.
43. [43] S. G. Clendenon et al., "A simple automated method for continuous fieldwise measurement of microvascular hemodynamics," *Microvascular research*, vol. 123, pp. 7–13, 2019.
44. P. Prentašić et al., "Segmentation of the foveal microvasculature using deep learning networks," *Journal of biomedical optics*, vol. 21, no. 7, p. 075 008, 2016.
45. R. Nivedha, M. Brinda, K. Suma, and B. Rao, "Classification of nailfold capillary images in patients with hypertension using non-linear svm," in *2016 International Conference on Circuits, Controls, Communications and Computing (I4C)*, IEEE, 2016, pp. 1–5.

46. W. S. Noble, "What is a support vector machine?" *Nature biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006.
47. K. Suma, K. Indira, and B. Rao, "Fuzzy logic based classification of nailfold capillary images in healthy, hypertensive and diabetic subjects," in 2017 International Conference on Computer Communication and Informatics (ICCCI), IEEE, 2017, pp. 1–5.
48. K. Suma, V. Sasi, and B. Rao, "A novel approach to classify nailfold capillary images in indian population using usb digital microscope," *International Journal of Biomedical and Clinical Engineering (IJBCE)*, vol. 7, no. 1, pp. 25–39, 2018.
49. P. Javia, A. Rana, N. Shapiro, and P. Shah, "Machine learning algorithms for classification of microcirculation images from septic and non-septic patients," in 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, 2018, pp. 607–611.
50. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
51. F. Ye, S. Yin, M. Li, Y. Li, and J. Zhong, "In-vivo fullfield measurement of microcirculatory blood flow velocity based on intelligent object identification," *Journal of biomedical optics*, vol. 25, no. 1, p. 016 003, 2020.
52. C. Ning, H. Zhou, Y. Song, and J. Tang, "Inception single shot multibox detector for object detection," in 2017 IEEE International Conference on Multimedia Expo Workshops (ICMEW), 2017, pp. 549–554. DOI: 10.1109/ICMEW.2017.8026312.
53. Y. S. Hariyani, H. Eom, and C. Park, "Da-capnet: Dual attention deep learning based on u-net for nailfold capillary segmentation," *IEEE Access*, vol. 8, pp. 10 543–10 553, 2020.
54. G. Dai et al., "Exploring the effect of hypertension on retinal microvasculature using deep learning on east asian population," *PloS one*, vol. 15, no. 3, e0230111, 2020.
55. H.-C. Jo, H. Jeong, J. Lee, K.-S. Na, and D.-Y. Kim, "Quantification of blood flow velocity in the human conjunctival microvessels using deep learning-based stabilization algorithm," *Sensors*, vol. 21, no. 9, p. 3224, 2021.
56. O. Oktay et al., "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.
57. G. Bradski and A. Kaehler, "Opencv," *Dr. Dobb's journal of software tools*, vol. 3, 2000.
58. R. Nishihara and P. Moritz, *Ray v2.0.0.dev0*. [Online]. Available: <https://docs.ray.io/en/master/whitepaper.html>.
59. P. KaewTraKulPong and R. Bowden, "An improved adaptive background mixture model for real-time tracking with shadow detection," in *Video-based surveillance systems*, Springer, 2002, pp. 135–144.
60. Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *IEEE signal processing magazine*, vol. 26, no. 1, pp. 98–117, 2009.
61. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
62. openCV, *Contour approximation method*. [Online]. Available: https://docs.opencv.org/3.4/d4/d73/tutorial_py_contours_begin.html.
63. V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ser. ICML'10, Haifa, Israel: Omnipress, 2010, pp. 807–814, ISBN: 9781605589077.
64. C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," *arXiv preprint arXiv:1811.03378*, 2018.
65. R. Nishihara and P. Moritz, *Ray v1.0.1*. [Online]. Available: <https://docs.ray.io/en/ray-1.0.1/serve/faq.html>.
66. K. Srinath, "Python—the fastest growing programming language," *International Research Journal of Engineering and Technology*, vol. 4, no. 12, pp. 354–357, 2017.
67. A. S. Saabith, M. Fareez, and T. Vinothraj, "Python current trend applications-an overview," *International Journal of Advance Engineering and Research Development*, vol. 6, no. 10, 2019.
68. S. Raschka, J. Patterson, and C. Nolet, "Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence," *Information*, vol. 11, no. 4, p. 193, 2020.
69. I. Stašćin and A. Jović, "An overview and comparison of free python libraries for data mining and big data analysis," in 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), IEEE, 2019, pp. 977–982.
70. G. Piatetsky, "Python leads the 11 top data science," *Machine Learning Platforms: Trends and Analysis*, 2019.
71. F. Pedregosa et al., "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.
72. C. A. Navarro, N. Hitschfeld-Kahler, and L. Mateu, "A survey on parallel computing and its applications in data-parallel problems using gpu architectures," *Communications in Computational Physics*, vol. 15, no. 2, pp. 285–329, 2014.



This page is intentionally left blank



GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY: C
SOFTWARE & DATA ENGINEERING
Volume 22 Issue 2 Version 1.0 Year 2022
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Steam Package Boiler Expert System for Control and Maintenance of Fertilizer Plants using Rule-Base Fuzzy Logic

By Aziboledia F. Boye, Nuka D. Nwiabu & Bennett E. Okoni

Rivers State University

Abstract- Generally, expert systems have been found very useful and even in fertilizer plants it has been deployed in handling operations in critical sections, such as material handling systems, online detection systems, granulation, air compressor among others. This paper presents research work for steam package boiler expert system for control and maintenance of fertilizer plants using rule-base fuzzy logic hybrid system, which has not been benefited much from expert system. The system handles cause of boiler failures in terms of controlling and maintaining the functional chemical components of the boiler drum and feed water parameters. validation on the system consistency, correctness, and its precision with six (6) steam package boiler parameters test value cases was conducted involving fourteen (14) fertilizer plant boiler domain partitioners. The boiler drum and feed water qualities with less or higher test value worst-cases validates the boiler system, showing each of the parameter's bar turns red, as displayed on the boiler's panel, while on test value best-cases, validates the system, displaying green on the boiler's panel bar as users entered the right value of parameters as design specification.

Keywords: expert system, rule-base system, fuzzy logic, steam package boilers, dynamic simulation.

GJCST-C Classification: DDC Code: 623.873 LCC Code: VM741



STEAMPACKAGEBOILEREXPERTSYSTEMFORCONTROLANDMAINTENANCEOFFERTILIZERPLANTSUSINGRULEBASEFUZZYLOGIC

Strictly as per the compliance and regulations of:



RESEARCH | DIVERSITY | ETHICS

© 2022. Aziboledia F. Boye, Nuka D. Nwiabu & Bennett E. Okoni. This research/review article is distributed under the terms of the Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0). You must give appropriate credit to authors and reference this article if parts of the article are reproduced in any manner. Applicable licensing terms are at <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

Steam Package Boiler Expert System for Control and Maintenance of Fertilizer Plants using Rule-Base Fuzzy Logic

Aziboledia F. Boye ^α, Nuka D. Nwiabu ^ο & Bennett E. Okoni ^ρ

Abstract- Generally, expert systems have been found very useful and even in fertilizer plants it has been deployed in handling operations in critical sections, such as material handling systems, online detection systems, granulation, air compressor among others. This paper presents research work for steam package boiler expert system for control and maintenance of fertilizer plants using rule-base fuzzy logic hybrid system, which has not been benefited much from expert system. The system handles cause of boiler failures in terms of controlling and maintaining the functional chemical components of the boiler drum and feed water parameters. validation on the system consistency, correctness, and its precision with six (6) steam package boiler parameters test value cases was conducted involving fourteen (14) fertilizer plant boiler domain partitioners. The boiler drum and feed water qualities with less or higher test value worst-cases validates the boiler system, showing each of the parameter's bar turns red, as displayed on the boiler's panel, while on test value best-cases, validates the system, displaying green on the boiler's panel bar as users entered the right value of parameters as design specification. The expert system prevents damaged and malfunctioning as control the alkalinity, prevent scaling, both mechanical & chemical corrosion, forming, correct pH values and then the conductivity which deals with the feed boiler water and monitored the level in the boiler drum using the industry parameters (pressure, temperature, level and flow).

Keywords: expert system, rule-base system, fuzzy logic, steam package boilers, dynamic simulation.

I. INTRODUCTION

The discovery and development of expert systems recorded since in the early 1970s until today. Expert systems are an offspring of the more general area of study known as artificial intelligence (AI) (Pandit, 2013). An expert system is computer program that contains a significant portion of the specialized knowledge of a human expert in a specific, narrow domain, and emulates the decision-making ability of the human expert in that area (Emmanuel & Adekunle, 2012). Expert systems are one area of artificial intelligence. Expert system is computer software had been designed for Simulates human expert in acts and activates , by the system's capability to find new facts

Author α ρ: Department of Computer Science, Rivers State University, Port Harcourt, Nigeria.
e-mails: bayelsaforprogress2022@gmail.com,
nwiabu.nuka@ust.edu.ng, bennet.okoni@ust.edu.ng

from available facts and render advice, to teach and execute intelligent tasks (Jabbar & Khan, 2015). Expert system answer to questions about specific problems by inference, including inference of human knowledge that is experts in the field. expert systems should be able to explain the reasoning process and conclusion for the final user contributed by Simin et al., (2013). In expert systems, the knowledge to develop the system is derived from human experts (hence the name) (Tripathi, 2011). According to (Pandit, 2013), one of the principal attractions of expert systems is that they enable computers to assist humans in many fields. These systems are rule-based systems are used to store and manipulate knowledge to interpret. A rule consists of two parts: condition (antecedent) part and conclusion (action, consequent) part, i.e.: IF (conditions) THEN (actions) Antecedent part of the rule describes the facts or conditions that must exist for the rule to fire. Consequent describes the facts that will be established, or the action that will be taken or conclusion that will be made. Information in a useful way. They are often used in artificial intelligence applications and research (Pandit, 2013). Knowledge is the key success in the performance of an expert system. Acquisition of the desired knowledge for the problem domain was obtained from intelligent human experts, said Adekunle et al., (2018). Expert Systems have been designed to assist in solving set of various problems, for instance, expert systems for designing and planning, classification & identification, prediction, Repairs, forecasting, discovery, diagnosis, monitoring control (Tan et al., 2016) etc. however, expert system for control and maintenance of steam package boiler is new as this was not benefitted and so specifically deployed in fertilizer plants.

II. LITERATURE REVIEW

Mrudula (2013) gives reasons on his experts review article, that expert system models the knowledge of the human expert, both in terms of content and structure. His conclusion remains that one of the principal attractions of expert systems is that they enable computers to assist humans in many fields, and that these systems are rule-based systems are used to store and manipulate knowledge to interpret. According

to Amit (2012), the impurities found in boiler water depend on the untreated feed water quality, the treatment process used and the boiler operating procedures.

a) *Expert System Theory*

Generally, the higher the boiler operating pressure, the greater will be the sensitivity to impurities. Therefore, the American Trainco, Inc. (2006), states operators (users) are responsible for control and maintaining the boiler in a safe and efficient manner with sound engineering practices and manufacturer's specified maintenance procedures. Mrudula (2013) describes different types of expert systems. They are rule-based expert system, fuzzy expert system, frame based expert system, and hybrid expert systems, many expert systems are in real use and in the future one is likely to see more expert systems packaged with domain knowledge being sold. Amit (2012) added that boiler devices comprise of feed water system, steam system and fuel system, and that industries worldwide are going through an increasing and stiff competition and increased automation of plants, like the Ovation expert system. Seán et al., (2014) suggested in their paper that an expert system performance could discover excessive levels of blow down in steam package boilers, and diagnose the reason, which could be related to feed water conductivity, or component issues, which becomes necessary and in line with the main area were this dissertation is providing solutions.

b) *Expert System Research on Maintenance*

Edgar et al., (2010), contributed by developing an expert system for fault diagnostics in condition-based maintenance and that the goal of the system is provided to the non-experts in industrial maintenance with a list of possible failure modes, decisions to be adopted. Affirmed by Nwaibu & Adeyanjo (2012), is the development of user centered design approach to situation awareness which if applied to system will certainly aid the operation of the system.

Cucu et al., (2009) examining the causes of equipment failure and malfunction been a central application field of expert systems or of data bases, in general, Chee et al.,(2016), exploits the enhancement of expert systems in engineering, and therefore, an expert system implies into the engineering and manufacturing process can provide greatest assistance for operational workers during performances and deal with critical and important tasks. Bretz (1990) discussed on a comprehensive expert system that assists in power plant boiler failure analysis and the maintenance planning subsystem automatically prepares daily repair schedules, a work estimation plan, and work specifications. Konstantin, et al., (2020) discussed that ES are very advantageous for steam turbine unit (STU) diagnosis.

c) *Expert System Research on Control*

Emerson Process Management Power and Water Solution Inc (2011), Ovation expert system was designed to help you achieve operational excellence and create a sustainable competitive advantage, with standard hardware platforms, operating systems, and network architectures.

Krishan (2016), proposed a fuzzy expert system, using fuzzy control design a simulation system of fuzzy logic controller for water tank level control by using simulation package which is Fuzzy Logic Toolbox and Simulink in MATLAB. Anabik & Denis (2012), on boiler temperature and water control, designing an expert system, using fuzzy logic on FLC based boiler control, the proposed method consists of two sections. first section is to develop a steam temperature monitoring and control system and the second section consists of water level control. For both sections Fuzzy Logic Control will be used. Celin et al., (2016) proposed modeling and control of a 200 MW power plant using the embedded fuzzy approach. The embedded fuzzy based boiler control consists of models, and the output of the network is the interpolation of the models' using memberships.

In line with power evaluation and control, a fuzzy based expert system named Nuclear Power Plant Operator Evaluation Expert System (NPPOEX) is one of the latest expert systems which shows the use of expert system techniques in the personal evaluation process, Yangping et al., (2011).

Harald & Georg (2015) proposed Expert Systems (ES)– Smart Solutions for Power Plants that can support operators and engineers in different areas. Monitoring systems like SR::EPOS enable continuous gathering and analyzing of data and thus can help to detect various problems very fast to avoid a reduction of efficiency or damages

Marco et al., (2003) and Tavira-Mondragón et al.,(2011) agreed that Control remaining the center of operations in the industries, Seán et al., (2014) said operator expertise is generally capable of reacting to faults and performing corrective measures, guidance using expert systems will improve operation and reduce downtime.

III. METHODOLOGY

According to Anne (2013), when conducting research or degree projects, methods and methodologies are essential to plan and steer the work to achieve proper, correct, and well-founded results, and on disciplines within computer sciences, the focus is often on the work with the research rather than methods.

To produce a good system that will meet functionality and performance for users in boiler control and maintenance for fertilizer plants, the researcher

brought in his computing and instrumentation engineering knowledge geared towards his working experience of more than ten (10) years alongside with the domain practitioners to understand more of the activities of the plant. To attain these task or objectives, is by integrating action research method and object-oriented development (OODA) for the modelling.

a) *Action Research Method*

According to Bryman and Bell (2011), action research is “an approach in which the action researcher and a client collaborate in the diagnosis of the problem and in the development of a solution based on the diagnosis. One of the main characteristic traits of action research relates to collaboration between researcher and member of organization to solve organizational problems. In line with the quantitative and qualitative research approach proposed by Kusumarasdyati, (2016), the method has advantages added Bryman & Bell (2011).

Both the numerical values for the quantitative and non-numerical values for the qualitative were all under technical, practical, and emancipatory action research. The numerical values are from the boiler drum recommended quality and feed water recommended quality, whereas the non-numerical entails the boiler problems and causes extracted from the O&M of fertilizer plant manual. Also, the researcher conducted one-on-one interview with fourteen (14) of the fertilizer plant boiler domain practitioners, carrying out dynamic simulations during the design of proposed system, changes were made based on their recommendations on the application on consistency, correctness, and its precision. The dataset numeric values are corresponding to the physical or chemical state of the boiler system, control measures such as shutdown, bidirectional or one directional flow, open on excess fluid, or close on excess fluid can be used to control the performance of the steam package boiler to keep it in a workable condition. The performance and health of the steam package boiler can be control, measured and monitored using process measuring, control and sensing devices such as pressure gauges, temperature sensor (transducers), flow and level gauges, actuating valve etc. which are incorporated in the proposed system designed. The idea is to take all the useful dataset from the boiler manual according to its performance and capability in different operation points to model better system for plant control and maintenance as stated by Michael (2016), same as applied in user centered design approach to situation awareness, Nwiabu & Adeyanju (2012).

b) *Objective-Oriented Development Approach*

Nabil & Govardhan (2012) describe object-oriented (OO) approach as a system development methodology that follows an iterative and incremental approach to systems development. In each increment or phase, the developers move through the activities of gathering requirements, analyzing the requirements, designing the system, implementing the design, and testing the system. Adetokunbo et al., (2013) added that object-oriented approach models its processes using objects. That is, the solution of problems can be seen as a set of objects or computations performed in the context of objects. Data and the processes that act on the data are encapsulated within every object. Each object's data (attributes or states) are the properties that relate to the object.

Nabil & Govardhan, (2012), added that object-oriented development approach (OODA) uses a set of diagramming techniques known as the Unified Modeling Language or UM and focuses on the three architectural views of a system: functional, static, and dynamic, makes use of iterative and incremental steps, it gives opportunity to manage changes as they occur to user requirements. So, it is more prone to user satisfaction. In this research work, the steam package boiler is object to be controlled and maintained. The Java programming language is suitable for the proposed system development because each module in the system is seen as classes, this gives the environment to develop a user interface for interaction with the system.

IV. EXPERT SYSTEM ARCHITECTURE

The expert system architecture of the steam package boiler rule-base system and fuzzy logic is shown in Fig1. It shows the encoded knowledge of the domain expert put into the system's database where the inference engine would make use of it to process and determine the working state and errors of the system.

When the system runs on the manual operation mode (MOM), the user (field operators/engineer) enters the correct parameter set points which enable a corresponding output from the input signal and expected value of the affected plant system as authorized by their plant shift-in-charge (SIC) or plant operations manager. In the expert system shell, data flow bi-directional to check values from plant (field instruments) and from field operator or process system engineer navigating the system.

The description of the system architecture as shown in Figure1 below, has the processes of both fuzzification and defuzzification, which plays major roles in the system architecture. The fuzzy logic is found in the part of the inference engine, where the software system converts rules to code and codes back to rules.

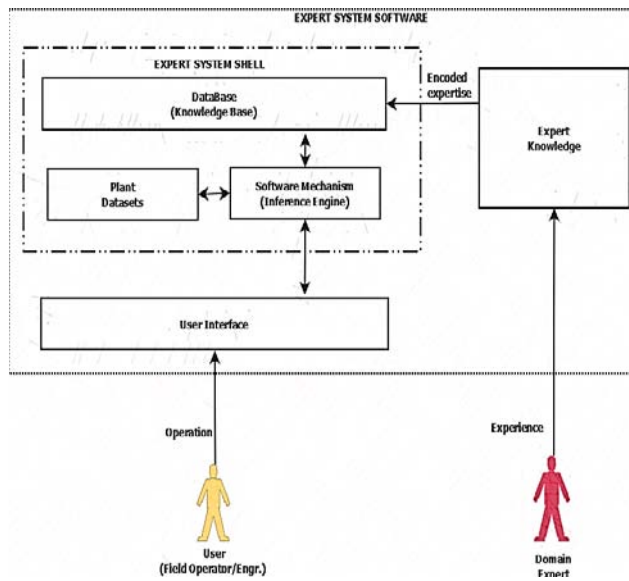


Figure 1: The Expert System Architecture

Domain Expert: The domain expert is the personnel either designing the ES shell or putting in his/her knowledge into the system. It has a ready and mostly accurate knowledge and experience on the proposed system shell to be designed. The expert then encodes their knowledge into the system, by making it understandable for the machine to read by fuzzification.

Knowledge Base (KB): The KB is the part of the ES shell which holds the encoded expert knowledge where the inference engine gets its own knowledge. It is knowledge warehouse or library of the entire ES shell.

Datasets/Software Mechanism (Rule of Engine): This is the process of converting a fuzzy system statement or expression into a software usable or understandable language. The fuzzified expression or language is in the form of a fuzzy system membership form. This is the process where the software makes use of the datasets in the system as encoded by the domain expert.

Plant Datasets: The plant datasets are then extracted from the inference engine (core of the fuzzy system) to be used for proper working of the plant. This process is bi-directional where the plant gets its datasets from the inference engine and gives its current datasets working condition to the inference engine for control.

User interface: this is the part of expert system architecture that helps the user to be in contact with the system so we can consider it as the main engine of expert system.

The User (Field Operators/Panel Engineers): The user on operations logs into the main system with an authenticated username and password. Users carried out required corrections of errors on the UI when the system is running on MOM and monitor and maintained the progress of the system on AMOs.

a) *Steam Package Boiler Rule-Base Fuzzy Logic*

The system has a inference kernel and a knowledge-based linked. The inference kernel (engine) is executed periodically to determine system output based on current system input. The knowledge-based contains membership functions and rules.

Fuzzification: The current input values are compared against stored input membership functions, usually in a program loop structure to determine the degree to which each linguistic variable of each system is true.

Defuzzification: Dissolves multiple degree ambiguous by putting raw fuzzy outputs into a composite numerical output.

Rule-Base Evaluation: on the rule-base evaluation, this processes a list of rules from the knowledge-based using current fuzzy input values to produce a list of fuzzy output linguistic variable.

Therefore, the system has the processes of both fuzzification and defuzzification, which plays major role in the proposed system architecture. The fuzzy logic is found in the part of the inference engine, where the software system converts rules to code and codes back to rules.

V. STEAM PACKAGE BOILER EXPERT SYSTEM DESIGN

a) *Datasets*

The datasets used are extracted from Section C, Boiler Manual (2020). Boiler drum water and Feed water recommended qualities datasets are used for the design of the fuzzy logic system. According to Amit (2012), the impurities found in boiler water depend on the untreated feed water quality, the treatment process used and the boiler operating procedures. As a rule, the

higher the boiler operating pressure, the greater will be the sensitivity to impurities. The boiler safety, emergency procedures datasets are used because that area is where fault solutions are placed. The datasets are placed in a tabular form as shown below in table 1 and 2.

Table 1: Boiler Drum Water Recommended Quality

Parameters	Unit	Value
Ph Value	-	10.8 - 11.4
Alkalinity as CaCO ₃	ppm	190
Sodium Phosphate as PO ₄	ppm	29 - 34
Total Dissolved solids	ppm	143
Silica SiO ₂	ppm	4
Oil and Organic matters	ppm	Nil

Table 2. Feed Water Recommendation Quality

Parameter	Unit	Value
General appearance		Clear and Colourless
Total hardness as CaCO ₃	ppm	Commercial zero
Total Fe	ppm	<0.025
Total Cu	ppm	<0.02
Oil and Organic	ppm	<0.1
Ph Value At 25°C	ppm	8.5 - 9.5
Total dissolved solid	ppm	<1.5
Reactive SiO ₂ max	ppm	<0.20

b) System Use Case Diagram

Use case diagram give simple overview of an interaction of the proposed system users can have more about what is happening. This detail can either be a simple textual, a structural description in a table or a sequence diagram and it is used to support requirement elicitation. This use case diagram shows the user's (field operator/engineer's) interaction with the proposed system, here the user navigates the proposed system and get the expected results/output.

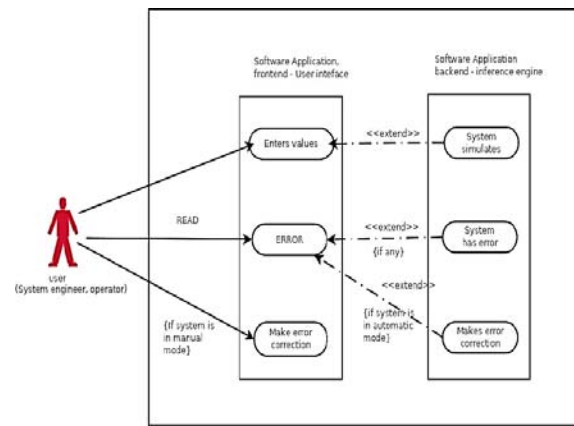


Figure 2: Use Case Diagram of the System

The use case diagram shows the actors (field operator/engineer) using the system. The user enters values required for the steam package boiler, to keeps the system on control, maintaining the required set points of pH values, feed water and boiler drum water level qualities etc., reduced downtime and boiler damages. When the system panel indicator displays red on either side of panels due to drifting of the steam boiler parameter set points, the user (field operator/engineer) should be able to see it and effect changes. On a MOM, the user makes more corrections during navigating the system, while the system on AOM, restores system errors on either boiler drum or feed water qualities as parameters drifted from their set points which effect system and make changes to keeps the boiler working efficiently.

c) System Activity Diagram

The activity diagram reveals the events of the user (field operator/Engineer) using the system. The user opens the app, check the system values. if the boiler runs with correct values as design specifications, the system runs smoothly and efficiently, otherwise, the user inputs the required parameter values to control and maintain the steam boiler system.

This system using the rule-base system & fuzzy logic technologies was designed to run in two modes, manual operation mode (MOM) and automatic operation mode (AOM). The auto run was selected due to faster and quicker response for equipment operations as instrument air is on FO/FC condition.



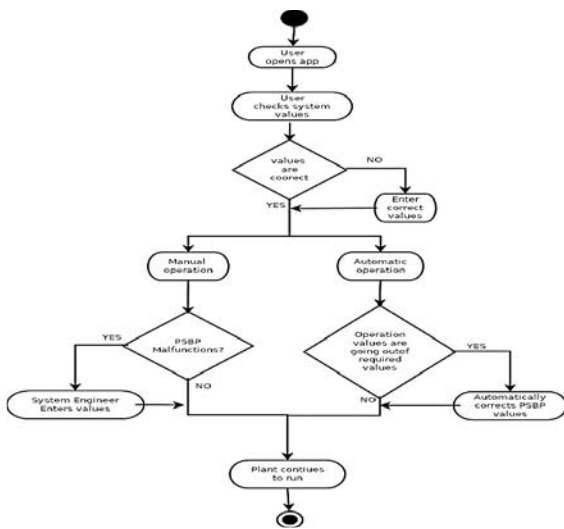


Figure 3: System Activity Diagram

On MOM, when the system malfunctioning is detected by a drift in setpoint parameters the user (field operator/engineer) inputs values for corrections which enables efficient control and smooth running of the steam boiler device, otherwise, the plant operation manager required for the maintenance team (Mech, Electrical and Instrument) to carry out maintenance by bye-passing required field equipment's. If the system is runs on AOM, and there is erratic or faults detections, and values are drifting from required default parameter setpoints, the system would automatically make corrections which aid continues running of boiler plant.

The process of designing this system, is aimed at providing detailed dataset from Table1 and Table2and other system elements that enable the development and implementation reliable. The system is designed is such a way that, the output variables or condition of the steam package boiler controls its input variables or condition of the boiler device. The pH value at 25oCin the table above has a range of value suitable for the application of fuzzy logic system. The FLS is designed as follows.

Algorithm

1. Define linguistic variables and terms (start)
2. Construct membership function for the terms (start)

Table 4: Membership function table -membership function replaced with actual numeric values. The actual rule from the inference engine will be as follows:

State	Target (s)				
-	Very/L	Low	Normal	High	Very/H
Very/L	No/C	8.75	9	9.25	9.5
Low	8.5	No/C	9	9.25	9.5
Normal	8.5	5.75	No/C	9.25	9.5
High	8.5	8.75	9	No/C	9.5
Very/H	8.5	8.75	9	9.25	No/C

3. Construct knowledge base of the rules (start)
4. Construct crisp data into fuzzy datasets using membership functions (Fuzzification).
5. Evaluate rules in the Rule Base (Inference Engine)
6. Combine results from each rule (Inference Engine)
7. Convert output data into non-fuzzy values (Defuzzification)

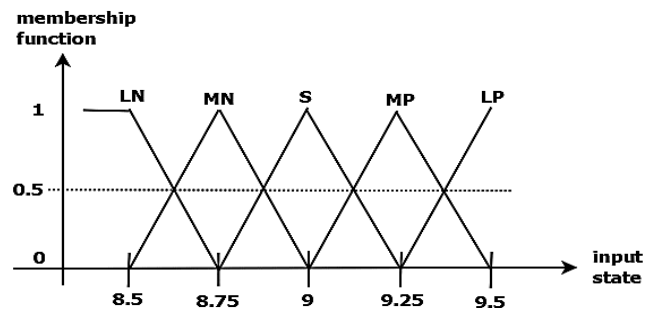
Development

1. State(s) = {very-low, low, normal, high, very-high}
2. Constructing membership function

Table 3: Ph Value at 25oc Showing Membership Values

M/Class	Class/R	Grade
LP	x = 9.5	L/value
MP	x = 9.25	H/Average
S	x = 9	Average
MN	x = 8.75	Lower/Average
LN	x = 8.5	Smallest/Value

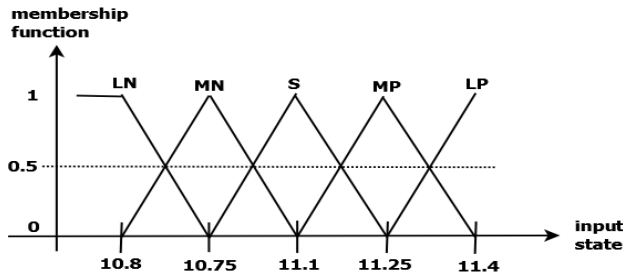
The above table3 has been used to design a membership function graph for the membership values in the table, as calculated with the datasets of pH value at 25oC ranging between8.5ppm (actual minimum value) to 9.5ppm (actual maximum value).



Graph 1: A graph of membership function of pH value at 25oC ranging between 8.5ppm (Mini value) to 9.5ppm (maxi value). 1and 4. Constructing a Knowledge Based Rule from the membership function.

From the boiler drum water recommended quantity and following the same steps of 1, the pH value (10.8 - 11.4) and the Sodium Phosphate as PO₄ (29 - 34) Fuzzy Logic Rule Base System will be developed using the parameters from table1.

The steps in the algorithm of 1aare followed in the system design and development of the pH value and Sodium Phosphate as PO₄ membership function and fuzzy logic rule base system.



Graph 2: Show above is a graph of the membership function of pH ranging between 10.8 ppm (min value) to 11.4 ppm (maxi value).

R1: IF STATE= (10.8 OR 10.75) AND TARGET = 11. THEN ALERT

R5: IF STATE= (10.8 OR 10.75 OR 11.1 OR 11.25) AND TARGET =11.4 THEN ALERT

Listing 5.a

R1: IF STATE= (10.8 OR 10.75) AND TARGET = 11.1 THEN SET STATE=11.1

R5: IF STATE= (10.8 OR 10.75 OR 11.1 OR 11.25) AND TARGET = 11.4

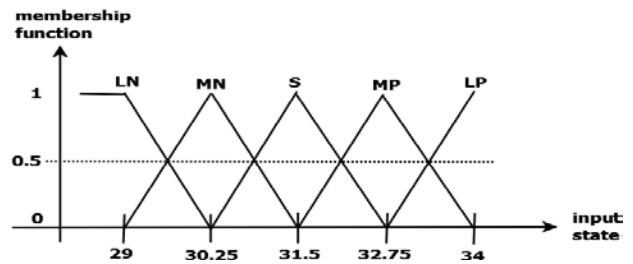
THEN SET STATE=11.1

Listing 5.b

Listing 5. Listing 5.a Manual mode, and Listing 5.b Automatic mode, are both the final Rule-Base System used in the Inference Engine of Sodium Phosphate as PO₄ ranging between 10.8ppm (mini value) to 11.5ppm (maxi value).

Table 5: Membership Function of Numeric Values of pH Ranging Between 10.8 ppm (mini value) to 11.4 ppm (maxi value) table.

State	Target(s)				
	V/Low	Low	Normal	High	V/High
-	No/C	10.75	11.1	11.25	11.4
V/Low	No/C	10.75	11.1	11.25	11.4
Low	10.8	No/C	11.1	11.25	11.4
Normal	10.8	10.75	No/C	11.25	11.4
High	10.8	10.75	11.1	No/C	11.4
V/High	10.8	10.75	11.1	11.25	No/C



Graph 3: A Graph of Membership Function of Sodium Phosphate as PO₄ ranging between 29ppm (minimum value) to 34ppm (maximum value).

Table 6: Membership Function of Numeric Values of Sodium Phosphate as PO₄ Ranging

between 29ppm (Min value) to 34ppm (Maxi value).

State	Target(s)				
	V/low	Low	Normal	High	V/high
-	No/C	30.25	30.5	32.75	34
V/Low	No/C	30.25	30.5	32.75	34
Low	29	No/C	30.5	32.75	34
Normal	29	30.25	No/C	32.75	34
High	29	30.25	30.5	No/C	34
V/High	29	30.25	30.5	32.75	No/C

R1: IF STATE= (29 OR 8.5) AND TARGET = 9 THEN ALERT

R5: IF STATE= (29 OR 8.75 OR 9 OR 9.25) AND TARGET =9.5 THEN ALERT

Listing 6.a

R1: IF STATE= (29 OR 30.25) AND TARGET = 30.5 THEN SET STATE=30.5

R5: IF STATE= (29 OR 30.35 OR 30.5 OR 32.75) AND TARGET =34

THEN SET STATE=30.5

Listing 6.b

Listing 6. Listing 4.a Manual mode, and Listing 4.b Automatic mode, are both the final Rule Base System used in the Inference Engine of pH value at 25oC ranging between 29ppm (mini value) to 34ppm (maxi value).

d) MembershipData and Fuzzy Logic System Class Diagram

The class diagram of the designed project class files is written in Java programming language (JPL). The diagram shows the private, public methods and variables of each class. It also shows how each class relates to one another, through inheritance (i.e., extension) or implementation (the use of interface class) as show by the lines. The expert system of the proposed system has two major selected class for the purposed of this research, which is the Membership Data class, and the Fuzzy Logic System are shown below.

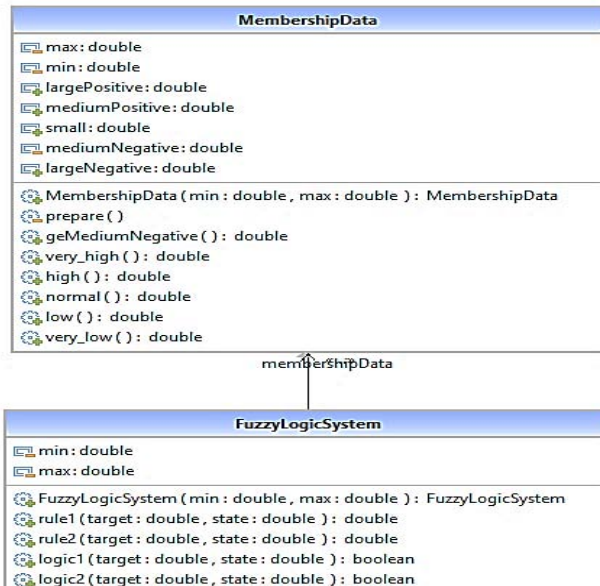


Figure 4: MembershipData and Fuzzy Logic System Class Diagram

e) System User Interface

The main purpose of this application user interface is to enhance the field operator/engineers with less effort to navigate and receive maximum desired outcome from the system. The steam package boiler ES Graphical user interface is menu-driven type and is shown in Fig7 below. Functions of the system were carried out with a dynamics simulation process in this research work.

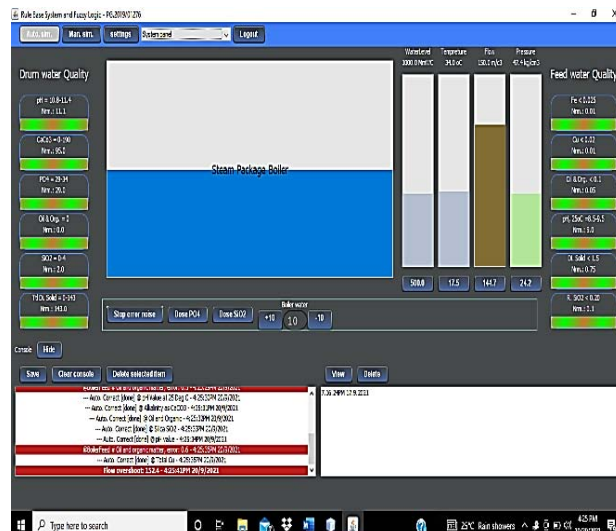


Figure 5: System User Interface

All manual and auto operations take place on this UI, it shows the window style view of the entire structure of the proposed system. Schematically it shows the drum water quality and feed water quality panels, steam package boiler drum, water level, temperature, flow, pressure, console etc. on its panel. Also, on the menu bar of the application shows the Auto sim button, manual sim button, setting button, system panel dropdown and logout menu button (Figure7).

f) Admin and User login Interface

This is the interface for both admin authorization and user login to have access to the main application.

The UI is designed to enable admin to provide his login details before the user (field operator/engineers) can have access to the main application. The Exit button is clicked to close the application. When credentials are correct the system gives access to the user to access the main application screen as shown in Figure 8.



Figure 6: Admin and User Login Interface

g) User Registration Window

On the registration window, the admin double clicking on the Expert System icon on the Desktop and splash up of the user access point window comprising of register and exit button.

The system is design for the admin to grant access to the user before he/she can have access to the system. An evaluated and validated window result for a successful registration process of a New User (Field operator/Process Engineer) carried out by the admin is shown in Figure7 below:

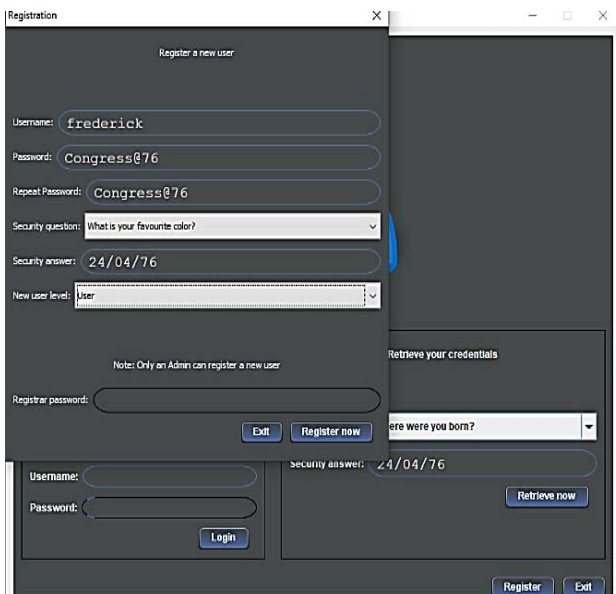


Figure 7: User Registration Window

Admin Authorization Window: This comprises of the following textbox, Admin password and authorize button, the (field operator/Engineer) username, password, and login button. **Retrieve your credentials Window:** it also comprises of the text boxes, security question, security answer and retrieve now button.

The above window interface also assist user with a forgotten login credentials, user (field operator/engineer) is expected to choose a security question and type the answer in the given text box and click the "Retrieve Now" button as shown in figure 6.

h) The Main System Control Panel Interfaces

The main control panel interface is where the users will navigate, monitor, controlled, and maintained the real-time conditions and readings (default parameters of the steam package boiler expert system with their operating setpoints for better efficiency, reliability, performance, and product throughput. The main control panel interface comprises of seven (7) panels as shown on Fig7. above.

- The Top Control Panel
- The Drum Water Quality Display Panel
- The Feed Water Quality Display Panel
- The Level, Temperature, Flow and Pressure Bars
- The Steam Package Drum
- The Bottom Control Panel

Each of these operational panels are describe below for better understanding.

The top control panel: This is menu situated on top of the main application comprises of simple "automatic and manual sim" buttons", "setting" button, panel selection drops down menu bar and "logout" button. The automatic simulation button is used to run and navigate the steam package boiler in an Auto Mode where all errors are corrected by the designed fuzzy logic system.

The Drum Water Quality Display Panel: This drum panel on the main application displays the real-time parameter values of the Total Hardness as CaCO₃ in ppm, TI (ppm Fe), TC (ppm Cu), Oil & Organic, pH value at 25 Deg. C, total dissolved solids (ppm TDS), and the Reactive Silica max (ppm SiO₂) of the water in the boiler drum, in their respective range of values. When there are low or high

values in the drum water quality parameters as calibrated setpoints desired for running of the steam package boiler ES, the green horizontal bars on the drum water panel turned red or green as shown fig8. Also, the boiler drum water quality should be continuously monitored and suitable adjustment in blowdown is carried out to maintain the drum water real-time values as per the steam package boiler design recommendations.

The Feed Water Quality Display Panel: This feed water panel on the main steam package boiler displays the real-time values of running plant with the setpoint of pH value, alkalinity of CaCO₃ in ppm, Sodium Phosphate as PO₄, total dissolved solids (ppm TDS), Reactive Silica max (ppm SiO₂), and Oil & Organic matters present in the feed as it goes into the steam package boiler drum, in their respective range of values. When there are low or high values in the feed water quality parameters as setpoints desired for the functional operation and running of the boiler system, error will be seen as the horizontal bar indicator turned red on the panel on the feed water quality panel.

The Steam Package Boiler Process Measurement Parameters

The four (4) bars on the main application interface represent these four main process measurements in the industry, first is the water Level in the drum (1000 MmWc), Temperature of the drum (34 Deg. C), Flow of the feed water (150 m/c3), and the Pressure that is built up in the drum as it is heated up (47.4 kg/cm³). These four (4) PMP can be seen on the main SPB main application interface as shown in Figure5. Each of the PMP values on the bar can be reset by mouse clicking the button below each of the parameter bars to get an input popup box. The user input textbox receives only the calibrated range specifies. Inputting right range values into the textbox will remove the red colour indicator on any of four parameter bars as required.

The Steam Package Boiler Drum: The steam package boiler is one the main fertilizer section that is monitored during operations. The boiler system drum can be seen on the main application interface. The water level is monitor with the result obtained from the four industry measurement parameters vertical bars representing pressure, temperature, level, and flow.

The Bottom Control Panel: This panel comprises of the "Stop Error Noise" used to stop any error noise during operations, the noise is coming due to system error, "Dose PO₄" and "Dose SiO₂" buttons are used for dosing as required to maintain feed & boiler water qualities at a desirable level, on clicking either of this button will bring a pop-up window with an input textbox which receives the value from any user (field operator/engineers) of the system.

The Console Panel: This panel is another important located on the bottom control. The result of using the console panel button is seen when it is mouse clicked, it comprises of the save, hide, view, and delete buttons. Each result of these buttons shown in fig10 below.

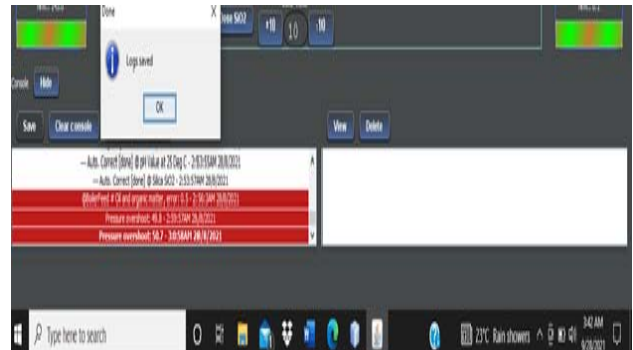


Figure 8: The Control Button

VI. EVALUATION AND RESULTS

The performance of every system after design and installation will be monitored to check if it's working properly, meaning to check if the system meets the design specifications. In other to evaluate, validate the correctness of proposed system by the boiler domain practitioners, the following parameters were used for dynamic simulation, Actual Minimum Range, Actual Maximum Range, Test values Worst-case and Test values Best-case were tested with the following conditions.

Actual Values: These values range are the boiler parameters from minimum to the maximum required limits of the system design specification. The healthy, proper, and effective performance on control and maintenance of the steam package boiler depends on these parameters operating within the required range and limits as shown below on the tabulated Table7 of worst-case and best-case test values.

Test Values: There are two type of test values used in this discussion, the worst-case and the best-case test values.

1. *Worst-Case:* The worst-case value are the values outside the expected functional range of the required plant parameters value. This worst-case value is used to monitor, examine, and evaluate the behaviour of the SPB if the water chemical component quality exceeds or falls below the required set value.
2. *Best Case:* The best-case values are the values which are set within the required functional range of the plant water chemical components.

Table 8: Parameters Test values (Worst-case/Best case) used for the dynamic simulation process to validate the system consistency, correctness, and its precision.

Parameters	AMR	AMR	Tv/Wc	Tv/Bc
BDWQ				
Ph Value	10.8	11.4	10	11.2
Po ₄ (Ppm)	29	34	43	32
Silica SiO ₂	0	4	6	2
FWQ				
Total Fe	0	0.025	2	0.010
Ph@25Deg C	8.5	9.5	8.2	8.9
TDS (ppm)	0	1.50	1.8	1.2

a) Dynamic Simulation of Operational Steam Boiler Drum Water Quality Parameters

The researcher explores a dynamic simulation process alongside fourteen (14) fertilizer boiler domain practitioners on three (3) qualities as displayed on the drum water panel of the boiler system comprising on its pH value, Sodium Phosphate (PO₄) and Silica max (SiO₂).



Figure 9: Drum Water Quality Display Panel

i. Dynamic Simulation for Worst-case and Best-case values for Drum Water Quality on pH value

The results of the dynamic simulation process of the parameter of pH value range of 10.8-11.4 with worst-case pH of 10. The system displaying this value, less than the designed setpoint (10.8), turns red the boiler parameter indicator bar and user clicking gives the system popup request to enter the right value to validate that the pH value of 10 is erratic. Also, the boiler panel displays green as the user enters best-case functional of pH value of 11.2 and this validates the system after evaluation.

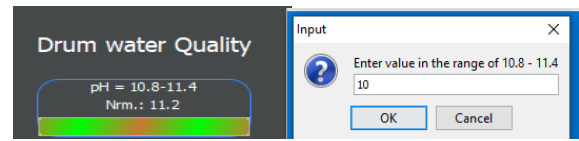


Figure 10: Ph Value of 10

ii. Dynamic Simulation for Worst-case and Best-case values for drum water quality on Sodium Phosphate as PO₄

On the dynamic simulation process of sodium phosphate as PO₄, the user views boiler panel displaying 43ppm of PO₄ as a worst-case value, verifying erratic value, and this turns boiler parameter indicator red. A popup request shows as user clicked on the indicator bar to enters best-case value of 32ppm, which turns the indicator of the PO₄parametergreen, validating the boiler operating in healthy condition. The PO₄systemdesign specification is between 29-32ppm.

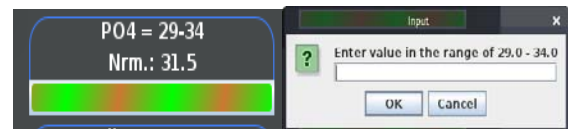


Figure 11: PO₄ Simulation Result

iii. Dynamic Simulations for Worst-case and Best-case values for Boiler Drum Water quality on SiO₂

The parameter design range of Silica max (SiO₂) value is 0.0-4.0 ppm. The boiler water quality must remain within its setpoint for the boiler to function efficiently as per design specifications. The results of the simulation with a worst-case of 6ppm, shows erratic display, verifying that SiO₂is high, showing red on the indicator bar. Its requires a user to right value, 2ppm of SiO₂, was entered, which turns parameter indicator bar green, validating the steam boiler running normally.

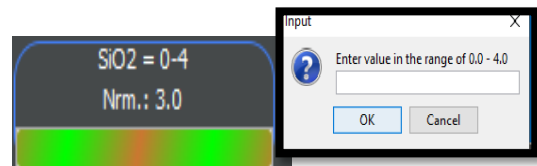


Figure 12: Silica Max Simulation Result

b) Dynamic Simulation for Operational Feed Water Quality Parameters

A dynamic simulation with test value worst-case and best-cases on feed water quality for iron (Fe) ppm, pH value at 25DegC and total dissolved solid (TDS) ppm were carried out and parameters displayed on the panel as evaluated and verified by fourteen (14) fertilizer boiler domain practitioners.



Figure 13: Feed Water Quality Display Panel

i. *Dynamic Simulation for Worst-Case and Best-Case Test Values on Iron Fe (Ppm)*

The iron (Fe) as a default parameter with range of 0.0-0.025ppm was simulated, and the feed water erratically displays worst-case of 2ppm a higher value of Fe, this turns the indicator parameter red. The erratic value was removed as a user clicking and system requesting right value to be entered. The user entered a test value best-case of 0.010 Fe (ppm), which is within the parameter setpoint range, and the parameter bar turns green, validating that the boiler device is enabled and working efficiently.

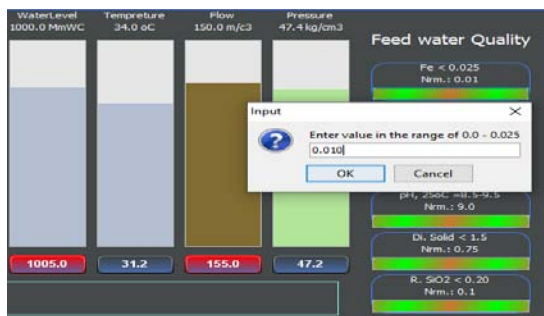


Figure 14: Fe (iron) Simulation Result

ii. *Dynamic Simulation for Worst-case and Best-case values on pH value at 25DegC*

The dynamic simulation results of the feed water quality for pH value at 25 DegC displaying pH value parameters of test value worst-case of 8.2-9.5 is shown below. The boiler system displaying a test value worst-case pH value of 8.2, less than the designed setpoint, turns indicator panel red, a user then clicked on the indicator bar and entered the right value range of pH value of say 8.9 to validate the system. The best-case value simulation which is a pH value of 8.9 @25DegC, turns indicator green, meaning value entered is within the boiler system designed specification.

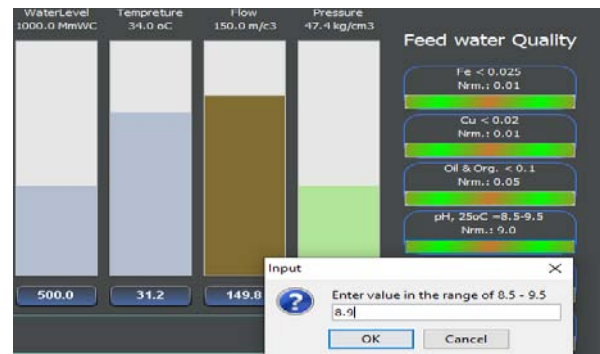


Figure 15: Ph Value at 25degC Simulation Result

iii. *Dynamic Simulation for Worst-Case and Best-Case Values for Feed Water Quality on (TDS) In Ppm*

Finally, a dynamic simulation for total dissolved solid (TDS) parameter was carried out with worst-case value of 1.80ppm and best-case value of 1.20ppm TDS. The feed water TDS panel indicator bar turns red at 1.80ppm which is erratic. A user is requested to enter correct value of the feed water quality at 1.20ppm which turns the TDS bar indicator green, this keeps boiler device in healthy operation.



Figure 16: TDS Simulation Result

c) *Analysis for Mean, Mean Absolute Deviation and Standard Deviation System Errors*

The researcher computes the mean, mean absolute deviation (MAD) and the standard deviation errors (SDE) of both manual and auto running with different time intervals to evaluate and verify which mode of the boiler system was preferable and the number of errors generated. The action research method deployed with quantitative observation of ten (10) consecutive run times of the boiler system on Table 8, comprising system running in manual and auto run. Each time corresponds with the number of errors generated.

Table 9: Proposed System Running on Manual and Auto Errors with Time (Minutes)

System Error Time (Minutes)	System Running on Manual	System Running on Auto
1	3	1
2	6	2
3	6	2
4	6	1
5	6	1
6	8	2
7	9	1
8	11	1
9	11	2
10	12	2

Table 10: Proposed Run System for Mean and Mean Deviation Error on Auto Run

Auto Run True System Error (x)	Mean Deviation Error on Auto Run x - μ
1	0.5
2	0.5
2	0.5
1	0.5
1	0.5
2	0.5
1	0.5
1	0.5
2	0.5
2	0.5
Mean (μ) E = $\frac{\sum x}{n} = 1.5$	(MAD) $\sum x - \mu = 5$

The mean (μ) error of the proposed system running on automatic mode is computed as follows:

$$\text{Mean } (\mu)\text{Error} = (\sum x)/n = (1+2+2+1+1+2+1+1+2+2)/10 = 1.5 \dots (1)$$

Therefore, the mean (μ) error of automatic system run mode is 1.5

The mean deviation error of the steam package boiler system running on auto mode is computed as follows:

$$\text{The Mean Deviation Error} = (\sum |x - \mu|)/n, \dots (2)$$

where x is the true mean error, μ is the mean error and n the sum of system errors in auto run of the proposed system.

$$\begin{aligned} \text{Standard Déviation} &= \sqrt{(\sum |x - \mu|^2)/n}, \dots (3) \\ &= \sqrt{(5^2/10)} = \sqrt{(25/10)} = 1.58113883 \approx 1.58 \end{aligned}$$

From Table10 of the error analysis, the auto run, minimum error is 1 in 1 minute while the maximum error is 2 in 10 minutes. The errors in minute are deduced from the log save error view button on the console and viewing the save logs screen.

On auto run mode, Error needed correction = overshoot + errors, this is subtracted from the errors correction done by the system, i.e., Correction Done-Needed correction = System errors, for instance in 2-minute, the system error gives 2 error on auto run. The overshoot was 8, error was 4, error needed correction is 12, correction was 14. This is the same with 4 minutes etc.

On manual run mode, the minimum error is 3 in 1 minute while the maximum error is 12 in 10 minutes analyzed from the log save error clicking view button on the console. Error margins were not considered during the MOM, since system runs efficiently on auto, whereas maintenance is done more on manual run of the

The Mean Absolute Deviation

$$\text{(MAD) Error} = (\sum |x - \mu|)/n = 5/10 = 0.5$$

In computing the standard deviation of the proposed system error on auto run, we must take the square root of the mean absolute deviation (MAD) squares it, divides by the total number of system errors (n) in auto run.

Table 11: Proposed Run System Mean and Mean Absolute Deviation Error on Manual Run

Manual True System Error (x)	Mean Deviation Error x - μ
3	4.8
6	1.8
6	1.8
6	1.8
6	1.8
8	0.2
9	1.2
11	3.2
11	3.2
12	4.2
Mean (μ) Error = $\frac{\sum x}{n} = 7.8$	(MADE) $\sum x - \mu = 2.4$

The mean deviation error of the proposed system running on manual mode is computed using the formula below:

$$\text{The Mean Deviation Error} = (\sum |x - \mu|) / n, \dots$$

Mean (μ) Error

$$= (\sum x) / n = (3+6+6+6+6+8+9+11+11+12) / 10 = 7.8$$

Therefore, the mean (μ) error of the proposed system running on manual mode is 7.8

The mean deviation error of the proposed system running on manual mode is computed using the formula below:

$$\text{The Mean Deviation Error} = (\sum |x - \mu|) / n,$$

where x is the true mean error, μ is the mean error and n the sum of system errors in auto run of the proposed system.

$$\begin{aligned} \text{The Mean Deviation Error (MAD)} \\ = (\sum |x - \mu|) / n = 24 / 10 = 2.4 \end{aligned}$$

In computing the standard deviation of the proposed system error on auto run, we must take the square root of the mean absolute deviation error (MADE) squares it, divide by the total number of system errors (n) in auto run.

$$\begin{aligned} \text{Standard Déviation} &= \sqrt{((\sum [|x - \mu|])^2) / n} = \\ &= \sqrt{(24^2) / 10} = \sqrt{576 / 10} = 57.6 \end{aligned}$$

Therefore, the SD of proposed system error is 57.6

where x is the true mean error, μ is the mean error and n, the sum of system errors in manual run of the proposed system.

The mean (μ) of the errors for the proposed system running on manual mode is computed as follows:

VII. CONCLUSION

This paper is focused on expert system for control and maintaining steam package boiler parameters of the feed water and boiler drum recommended qualities using a hybrid technology of rule-base system and fuzzy logic. The dynamics simulation performs showed relatively high performance of consistency, correctness, and its precision of the system. The system can be modified, used as a platform by boiler trainers, especially for new intakes in all boilers operated allied industries.

The design system will prevent damaged and malfunctioning as control the alkalinity, prevent scaling, both mechanical & chemical corrosion, forming, correct pH values and then the conductivity which deals with the feed boiler water. Acceptable error margined were not considered during this research work.

Furthermore, we discovered in this paper that from 1 to 10minutesrun time for AUTO RUN form figure17 gives 2.8% errors as compared to the 1 to 10 minutes time interval on MANUAL RUN that gives 80.2% error. This speaks more for users less effort in user interface application usage on AUTO operations better than MANUAL operations, leading with a recommendation that AUTOMATION in industrial plant processes (fertilizer plant) for control and maintenance make it more consistence, correct and precise than MANUAL process of operations.

REFERENCES RÉFÉRENCES REFERENCIAS

1. Amit, K. J. (2012). An Approach towards Efficient Operation of Boilers. International Journal of Scientific & Engineering Research, 3(6).
2. Adetokunbo et al, (2012), Software Engineering Methodologies: A Review of the Waterfall Model and Object-Oriented Approach, International Journal of Scientific & Engineering Research, 4(7).
3. American Trainco Inc. (2006). Boiler Operation Maintenance & Safety Study Guide American Trainco Real World Training, American Trainco, In. www.AmericanTrainco.com
4. Anne Håkansson (2013), Portal of Research Methods and Methodologies for Research Projects and Degree Projects, WORLDCOMP'13 - The 2013 World Congress in Computer Science, Computer Engineering, and Applied Computing, 22-25 July

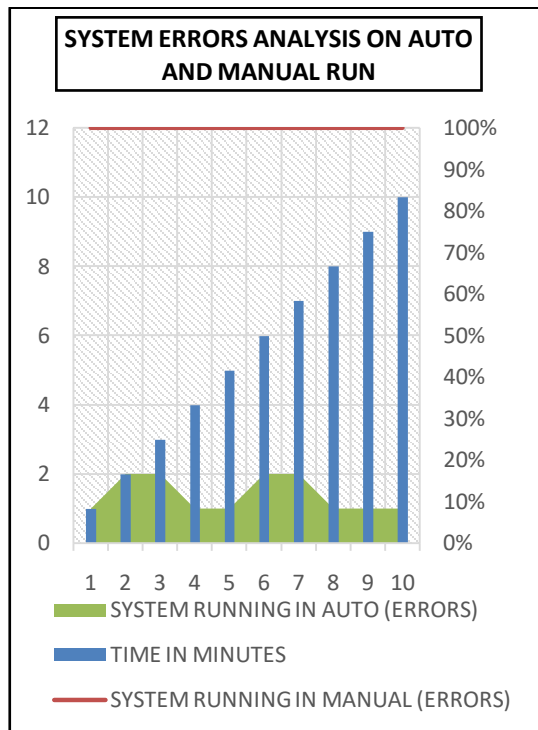


Figure 17: System Error Graphical Analysis on Manual and Auto Run

- 2013, FECS'13 - The 2013 International Conference on Frontiers in Education: Computer Science and Computer Engineering
5. Adefemi A. Adekunle, Peter P. Ikubanni² & Olayinka O. Agboola (2018), An Expert System for Automobile Repairs and Maintenance, *Mindanao Journal of Science and Technology*, Vol.16 (2018) 41-5
 6. Anabik, S. & Ashok, S. D. (2012). Fuzzy Logic Approach for Boiler temperature and water level. *International Journal of Scientific and Engineering Research*, 3(6).
 7. Asabere, N. Y. & Kusi-Sarpong, S. (2012). A Mobile Vehicle Expert System for the Automobile Industry. *International Journal of Engineering Research and Applications*, 2(6), 1108-1123.
 8. Bretz, E.A. (1990), Expert System Enhance Decision Making Abilities of O&M Permission Electrical world, 39-48.
 9. Celin S, Anne Frank Joe, Rajalakshmi G, Thaj Mary Delsy T, Jamuna Rani D (2016), Embedded Fuzzy Based Boiler Control, *International Journal of Robotics and Automation(IJRA)*, 5(2).
 10. Cucu I, Codruța D, and Imola D. (2009), Using Expert Systems in The Management of Industrial Equipment Maintenance, *Annals of the University of Petroșani, Economics*, 9(1), 2009, 191-198
 11. Castantio, M. J. P., Hermandes, F. Dedres, A. M., Rantenberge, S. & Billis, A. (2013). Expert Systems with Applications, Fuzzy expert system for predicting pathological stages of prostate cancer, 40, 466-470
 12. Emmanuel C. Ogu, Adekunle, Y.A. (2013), Basic Concepts of Expert System Shells and an Efficient Model for Knowledge Acquisition, *International Journal of Science and Research (IJSR)*, India Online ISSN: 2319-7064, Vol. 2 (4).
 13. Edgar Amaya Simeón & Ricardo Ribeiro Gudwin, (2010), An Expert System for Fault Diagnostics in Condition Based Maintenance, *ABCM Symposium Series in Mechatronics - Vol. 4 - pp.304-313*.
 14. Tripathi, K. P. (2011), "A Review on Knowledge-based Expert System: Concept and Architecture," *Artif. Intell. Tech. - Nov. Approaches Pract. Appl.*, vol. 4, no. 4, pp. 19–23.
 15. JabbarH. K. &R. Z.; Khan, "Survey on development of expert system in the areas of Medical, Education, Automobile and Agriculture," *Comput. Sustain. Glob. Dev. (INDIACom)*, 2015 2nd Int. Conf., vol., no., pp.776- 780, 11-13 March 2015
 16. Krishan K. (2016), Fuzzy Logic Execution in Boiler Control, *IJSRD - International Journal for Scientific Research & Development* | Vol. 4, Issue 05, 2016 | ISSN (online): 2321-0613.
 17. Konstantin E. Aronson, Boris E. Murmanskii, Ilia B. Murmanskii & Yuri M. Brodov, (2020), An Expert System for Diagnostics and Estimation of Steam Turbine Components' Condition, *Int. J. of Energy Prod. & Mgmt.*, Vol. 5, No. 1 (2020) 70-81
 18. Haider, K. J. & Rafiqul, Z. K. (2016). Tools of Development of Expert Systems: A comparative study, *Proceedings of the 10th INDIACom; INDIACom-2016; IEEE Conference ID: 37465 2016 3rd International Conference on "Computing for Sustainable Global Development"*, 16th - 18th March, 2016 Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA).
 19. Mrudula, P. (2013). Expert System-A Review Article, *International Journal of Engineering Sciences & Research Technology*, 2(6).
 20. Seán, H., Ken, B. & Dominic, O, (2014). Application of Expert Systems to Industrial Utility Equipment Optimisation, *Proceedings of the 14th International Conference for Enhanced Building Operations*, Beijing, China, September 14-17.
 21. Simin, S. S., Fatemeh, M., Fatemeh, A., Marjan, T. & Afsaneh, A. (2013). Investigate the Effect of Expert Systems Application on Management Performance. *Institute of Interdisciplinary Business Research. Interdisciplinary Journal of Contemporary Research in Business*, 478 4(1).
 22. Chee T. et al., (2016), The Application of Expert System: A Review of Research and Applications, *ARPN Journal of Engineering and Applied Sciences*, ISSN 1819-6608, VOL. 11, NO. 4, February 2016.
 23. Jarraya, Y., Souhire, B., Adel, M. A. & Ajith, A. (2013). Fuzzy Modelling System Based on Hybrid Evolutionary Approach, *ResearchGate*.
 24. Krishan, K. (2016). Fuzzy Logic Execution in Boiler Control, *IJSRD - International Journal for Scientific Research & Development*, 4(05), 2321-0613.
 25. Michael, A. B. B. (2016). Model Predictive Fuzzy Control of a Steam Boiler, Master's degree in Automàtic and Robotics. *Escola Tècnica Superior d'Enginyeria Industrial de Barcelona*.
 26. Muhammad, A., Suleman, A., Humza, W. & Nuka Nwiabu and Ibrahim Adeyanju (2012), User Centered Approach to Situation Awareness, *International Journal of Computer Applications*, (0975-8887), Vol.49-No17, July 2012.
 27. Yunusoglua, M. G. & Selim, H. (2013). Expert System with Applications, A fuzzy rule-base expert system for stores evaluation and portfolio construction: Approach to Istanbul Stock Exchange, 40(3), 908-920.
 28. Yangping Zhou, Xiang Fang and Xu Hong He (2011). "Use of an expert system in a personnel evaluation process". *Proceedings of IEEE International Conference Quality and Reliability*, Beijing. 2011: 15-19

This page is intentionally left blank



Integration of the Big Data Environment in a Financial Sector Entity to Optimize Products, Services and Decision-Making

By Ulises Roman Concha, José Huapaya Vásquez, Guillermo Morales Romero
& Dominga Cano Ccoa

National University of Juliaca

Abstract- This article describes the integration from big data environment in the management of products and services from a banking entity with optimizing financial products and decision-making. Actually, there are many financial entities where their different business areas have isolated databases, causing greater consumption of computer resources, maintainability and, in many cases, process delays. This problem becomes critical specially if there is a transnational company because data needs can vary geographically despite being the same functional area. The Data Architecture area proposed guidelines such as centralizing information in a big data environment, ensuring progressive accessibility from users for new financial analytics initiatives and thereby reducing isolated data. The agile, Scrum framework supported the advanced analytics pilot which comprising developments in the data ingestion layer (data lake) through the distributed processing from Apache Spark; and information consumption through Sandboxes, which one, users performing the analysis, visualization and prediction from data.

Keywords: big data, banking entity, financial sector, decisión-making, sandbox, data lake, scrum, spark.

GJCST-C Classification: DDC Code: 005.7 LCC Code: QA76.9.B45



INTEGRATION OF THE BIG DATA ENVIRONMENT IN A FINANCIAL SECTOR ENTITY TO OPTIMIZE PRODUCTS SERVICES AND DECISION MAKING

Strictly as per the compliance and regulations of:



RESEARCH | DIVERSITY | ETHICS

Integration of the Big Data Environment in a Financial Sector Entity to Optimize Products, Services and Decision-Making

Ulises Roman Concha ^α, José Huapaya Vásquez ^σ, Guillermo Morales Romero ^ρ & Dominga Cano Ccoa ^ω

Abstract- This article describes the integration from big data environment in the management of products and services from a banking entity with optimizing financial products and decision-making. Actually, there are many financial entities where their different business areas have isolated databases, causing greater consumption of computer resources, maintainability and, in many cases, process delays. This problem becomes critical specially if there is a transnational company because data needs can vary geographically despite being the same functional area. The Data Architecture area proposed guidelines such as centralizing information in a big data environment, ensuring progressive accessibility from users for new financial analytics initiatives and thereby reducing isolated data. The agile, Scrum framework supported the advanced analytics pilot which comprising developments in the data ingestion layer (data lake) through the distributed processing from Apache Spark; and information consumption through Sandboxes, which one, users performing the analysis, visualization and prediction from data. All this framed in stages such as: Geographical Diagnosis, Platform Validation, Design and Development of the Pilot.

Keywords: big data, banking entity, financial sector, decisión-making, sandbox, data lake, scrum, spark.

I. INTRODUCTION

The integration from a Big Data environment in a financial institution requires management concepts, experience in developing software, data modeling and agile frameworks.

The limitations discovered in a financial institution in Peru were based that the architectural pieces were installed on servers in Mexico, causing dependence on operators in that country for the administration from the Big Data environment. In addition to this, the users have too much operability rooted in their legacy systems which leads them to be resistant to change, so, in this manner, the data architecture area must provide information and knowledge to the users so that they get involved and

feel satisfied with the improvements from a Big Data environment

The integration from this Big Data environment was applied using the agile SCRUM framework, with a multidisciplinary team bringing to the own business users such as: Data Scientists hand in hand with the developers and data architects.

The solutions that are performed on this BIG DATA environment are based on the advantages of Apache Spark as a processing engine and HDFS (Hadoop Distributed File System) as a storage tool.

The analytical environments for Data Scientists which are Sandboxes, which are workspace with the possibility of running notebooks with the development in Python and Scala allowing Data Scientists to performing Machine Learning initiatives or work analytical models with different distributed processing libraries, these Sandboxes have access to consume certified and governed data hosted on the Data Lake, which is the storage component of the environment.

The Data Lake contains the raw data which is storage from the different applications from the business areas; All of this raw information is mastered to be consumed by users, which means that through treatment from data, the information is collected to functional concepts called "application units", with this, these mastered data achieve value and veracity to be exploited by the different business areas

Being the main stages of the integration from the Big Data environment: Geographic Diagnosis, Platform Validation, Design and Development of the Pilot.

II. METHODS

The implementation of the integration from the BIG DATA environment began with a pilot in the Business Development area (commercial area of the bank's products); as detailed in the following stages, first determining the scope of the pilot, going through the validation of tools to the deployment of products which validate the integration of the environment and makes it available for other analytical projects.

a) Geographic Diagnosis

The implementation area adopted is Business Development (or commercial area) for having several

Author α: Professor of the Department of Computer Science, UNMSM, LIMA-PERU. e-mail: nromanc@unmsm.edu.pe

Author σ: Graduated from the Faculty of Systems Engineering and Informatics, LIMA-PERU. e-mail: jose.alberto.huapaya@gmail.com

Author ρ: Professor at the National University of Education, LIMA-PERU. e-mail: dm.cano@unaj.edu.pe

Author ω: Professor at the National University of Juliaca, LIMA-PERU. e-mail: gmorales@une.edu.pe

analytical models for generating financial products and for interacting with other business areas which are linked with the Finance and Risks areas, and a multidisciplinary team was formed with different roles such as: Data Scientist, Data Developers, Data Architects, Project Managers and Agile Coaches.

Taking initiatives in a commercial area has many benefits of problem solving because we have advanced analytics possibilities in the Customer Network, Prediction of abandonment, geolocation, fraud detection, etc. (Galeano et al., 2019)

The chosen pilot consisted of migrating the entire monthly process from credit card campaigns for natural persons which until that moment took around 35 days from its conception to the arrival of the credit card offer to each client.

Taking advantage of the distributed processing and storage of a Big Data platform, the challenge is to reduce this process to 5 days.

b) Formation of Critical Mass

We started training with programming languages such as: Scala and Spark for the processing the Data Scientist will migrate and with the handling of HDFS. As well as working in a Sandbox with Notebooks where it can be developed in these languages on the Big Data environment.

For the developer roles, which are training in the knowledge of HDFS and data ingestion with libraries provided by the global architecture area. This library is based on Scala and Spark and its objective is to give structure to the inputs (which are usually flat files or csv), which structures are used for the government of the Data Lake.

In addition to the structure, these ingest libraries provide basic and configurable data transformations, which are nothing more than DMLs (Data Manipulation Language) in Spark SQL, which would be the well-known Join, GroupBy, Union, etc. Also, tools with Continuous Integration for the deployment of these

intakes, using components such as Jenkins (to orchestrate the deployment pipeline) and Bitbucket (as a repository for configurations and jobs).

Lastly, agile coaches trained in agile frameworks which having a multidisciplinary team taking account Scrum ceremonies such as: Planning, Dailys, Reviews, Retrospective. From the Extreme Programming methodology is collected the use of user stories. The team needed to know the methodological framework to work together and in an incrementally manner because different management involved (Risks, Finances, Product Development and Engineering) were waiting deliverables every fourteen days for contributing resources. Scrum helping in the developing of iterative deliverables as part of the final product, through Sprints and stakeholder collaboration.

c) Platform Validation

Our financial entity has centralized the infrastructure for all of Latin America, reason for which the data architects from each country are responsible for the availability pieces. These pieces go through a reception process through POC (Proofs of Concept) where basic cases are executed, individually testing each piece with the aim of knowing the functionalities and scope before of showing it to the users who will require training. There is no necessary from stress testing or performance because it is required by the Quality Assurance Global team at the installation. From a piece. From in this manner, the Local Data Architecture team is the second functional validation front for the Big Data environment because there is a continuous communication with Global Architects.

These pieces are collected into two fronts like: Data Lake and Sandbox:

- *Data Lake:* All the pieces that are involved in the provisioning or ingesting of data governed to the Data Lake.

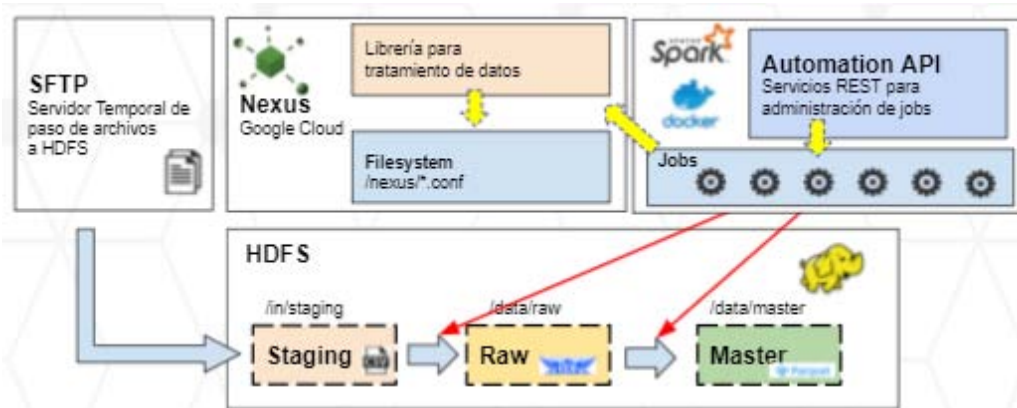


Figure 1: Distribution from the Infraestructure from Data Lake

- *SFTP*: Here it is observed which the transmission of input data is through the SFTP transfer protocol, this transmission is done through Control M which is a Batch programming tool, this tool serves the entire data lake provisioning life cycle.
- *HDFS*: In this table we have 3 layers of provisioning
 - *Staging*: Temporary storage layer where the files reside since its origin, for pilot purposes it was homologated to csv files. The temporality of these files should be the minimum possible because the information can contain the sensitive data (in the following layers, there is the possibility of encryption) and it should also be temporary because the flat files take up a lot of space that

must be used by other consumable formats. by users in Big Data.

- *Raw*: It is the raw data layer in the Big Data environment, this layer contains compressed files at the row level (AVRO format) for subsequent transformations, but not for queries and advanced analytics. This layer also serves as historical storage for its disk savings and also serves as a backup against possible reprocesses of the next layer.
- *Master*: Final layer of the Data Lake for which it is compressed at the columnar level with the aim to support exploratory analysis and processing of analytical models, this file format is PARQUET. This layer is the one that Data Scientists have access to in their respective Sandboxes.

Figure 2: Comparison of Formats of Files

File format	Schema integration	Compression support
Text/CSV (Shafranovich 2005)	–	–
JSON (Bray 2014)	+	–
Avro (Apache 2009a)	+	+
SequenceFile (Apache 2009b)	–	+
RcFile (He et al. 2011)	–	+
ORC file (Apache 2017)	–	+
Parquet (Apache 2013)	+	+

Source: (Plase et al., 2017)

As it's well-known in the previous image, Avro and Parquet stand out among the other file formats considering the integration of data structures and compression support; these characteristics are consistent with the manipulation of data in a Big Data environment.

- *Nexus Repository*: It is the repository of artifacts where store all the configuration files used in the feeding of files from the different layers of the Data Lake, as well as the libraries that are consumed in complex processing.

Actually, the replacement of this component with JFrog Artifactory, an Open Source tool with the same functionality as Nexus but that complies more efficiently with the installed ecosystem.

- *Automation Api*: It is the processing engine and clustering based on the Spark-based. It is distributed by a series of containers (Docker) and agents (Nodes) that distribute the cores efficiently to execute the different jobs that provision files in the Data Lake. As can be seen in Figure 1, in this layer storages the jobs (which will be used for information processing, such as feeding for the Data Lake), which are created and executed through rest

requests (communication between servers via HTTP).



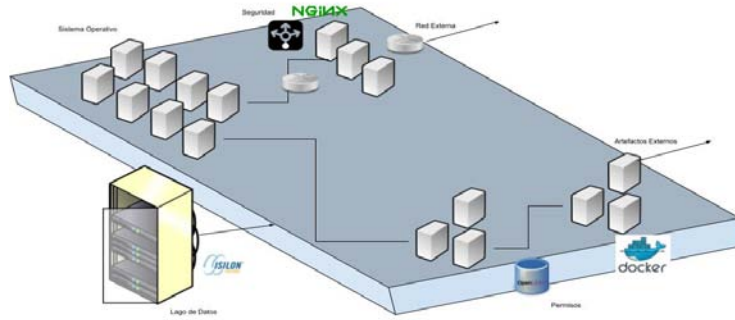


Figure 3: Diagram from Compounds with Tools from Open Source

- **Sandboxes:** The analytical components that consume the information from the Data Lake, described in the previous points, are distributed at the level of sandboxes which were mounted on

Google Cloud. In the actualization, we do not use Google services and the Jupyter Lab tools are mounted on the same servers of the financial institution.

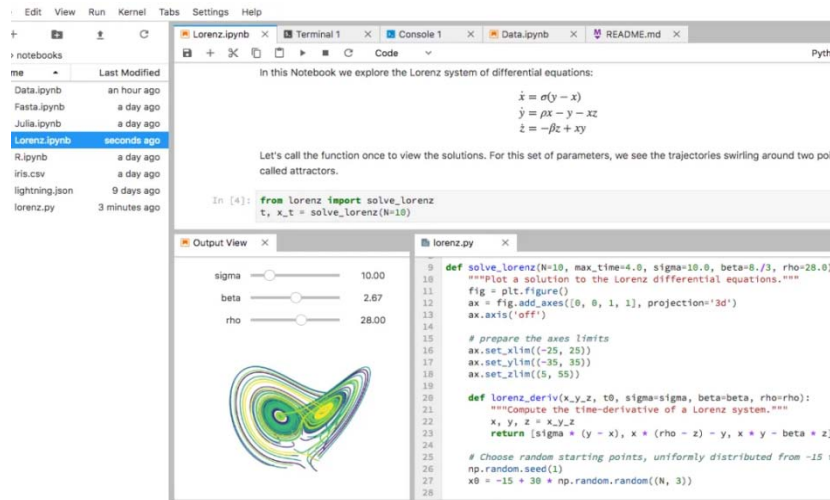


Figure 4: Jupyter and Advanced Analytics

Sandboxes contain specific permissions for the files in the Data Lake, according to the responsible for the sandbox, which means different Data Scientists can access to different sandboxes which can access different governed files in the Data Lake.

PySpark and ScalaToore (Python and Scala, respectively) and the access and profiling modalities for the Data Scientists to develop their processing and / or models.

For the campaign pilot, we created a sandbox and performing the basic processing libraries in

To finish the component validation, we integrate all the pieces described above in the Component Diagram of figure 5

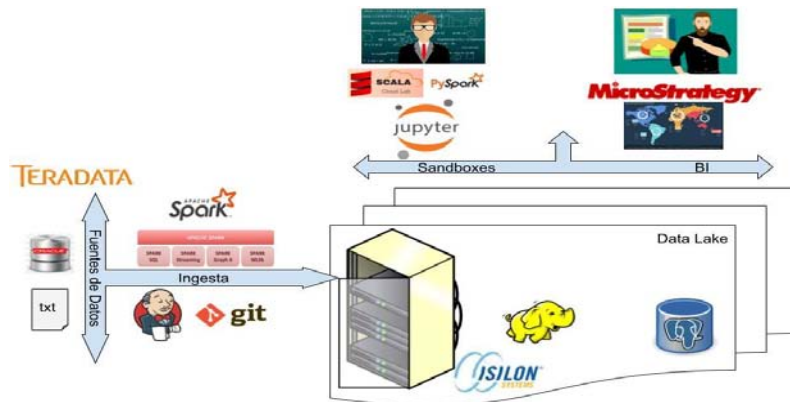


Figure 5: Components Diagram

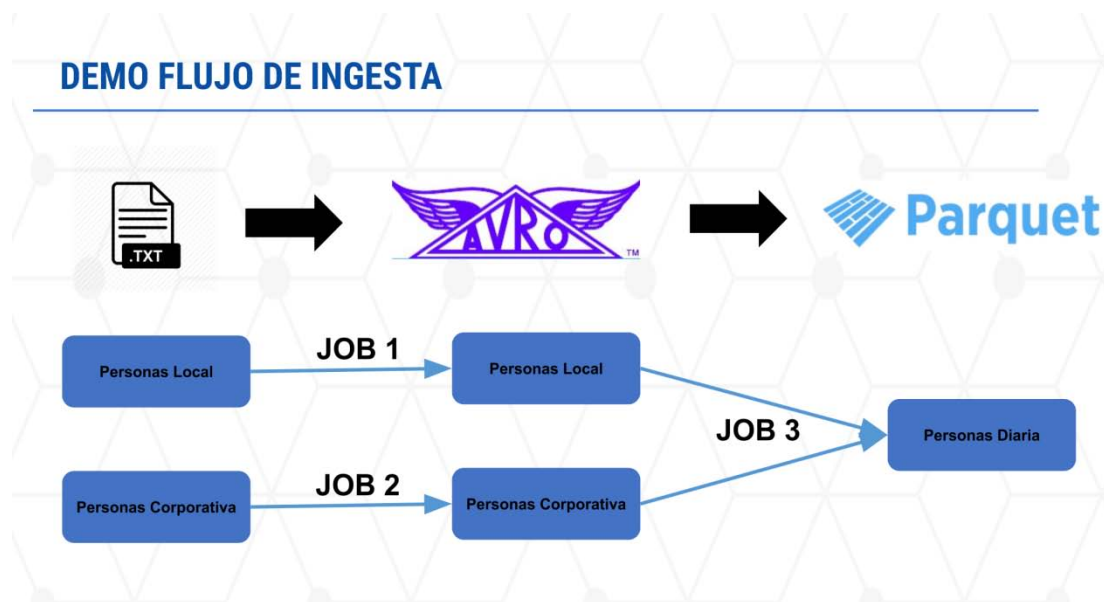
The components in Figure 5 follow the life cycle from the data in the financial institution's big data environment.

We start with the feed from the data from data sources such as: Teradata or flat files, we continue with the feeding process with the libraries described in the previous point, there are cases in which the feeding is complex due to business rules for which are developed specific procedures in spark, this being consider the third box, the central component and the most important is the Data Lake that currently contains HDFS, later other databases will be added (it is dotted in red), which continues are the neuralgic components from the Big Data environment: Sandboxes(for the development from advanced analytics or processing engines) and Business Intelligence tools (such as MicroStrategy) to design departmental reports to support decision-making, finally, the big data environment can serve as input for other applications, therefore, there is the scheme of "BD as a services" for which this component only replying to leaving the resulting files in a staging zone for the output from the environment so that it can be transmitted to different applications through Connect: Direct (IBM's secure transfer network)

d) *Design and Development of the Pilot*

At this stage, the developers with the tools already validated, proceeded to perform two activities in supplying the Data Lake to implement the Credit Card Campaigns (TC) pilot.

1. *Dictionary of Sources:* Analyze with the different bussiness areas, the source from the supplies for the TC campaign process. Once identified, they proceed to map the physical source from the table or file, describing attribute by attribute, specifying the correct data type and designing a relational data structure with global standards for registering this mapping in the data dictionary.
2. *Data Feeding:* Through the use from the ingesting libraries, previously validated, the dictionary files are loaded into the STAGING area (Flat Files), then to the RAW area (in Avro format) and finally to the MASTER area (in Parquet format). Each of these loading processes are nothing more than ETLs (Extraction, Transformation and Loading) that are elaborated in Business Intelligence.



Figur 6: Flow of Ingesting

The Jobs with which each executed ingest are mounted in Automation Api. This development environment is called: "Work" and there is a production replica environment which are called: "Live".

While the developers continue to provide the Data Lake, the Data Scientists work their models with test data in the available Sandbox, it should be noted that the Sandbox consumes information from the data lake, but it can also consume information which the own user uploads, but the latter as it is not governed, it only exists in the Data Scientist workspace, and not in the

Data Lake. This set of criteria is found in good intranet practices.

When the developers finish to provide the Data Lake, the Data Scientists will change their notebooks (Data Scientist development environment for working analytical models and procedures) to consume the governed sources (no longer the ones loaded by themselves); here begins the implementation of the pilot where business users are autonomous in said executions and certifications.

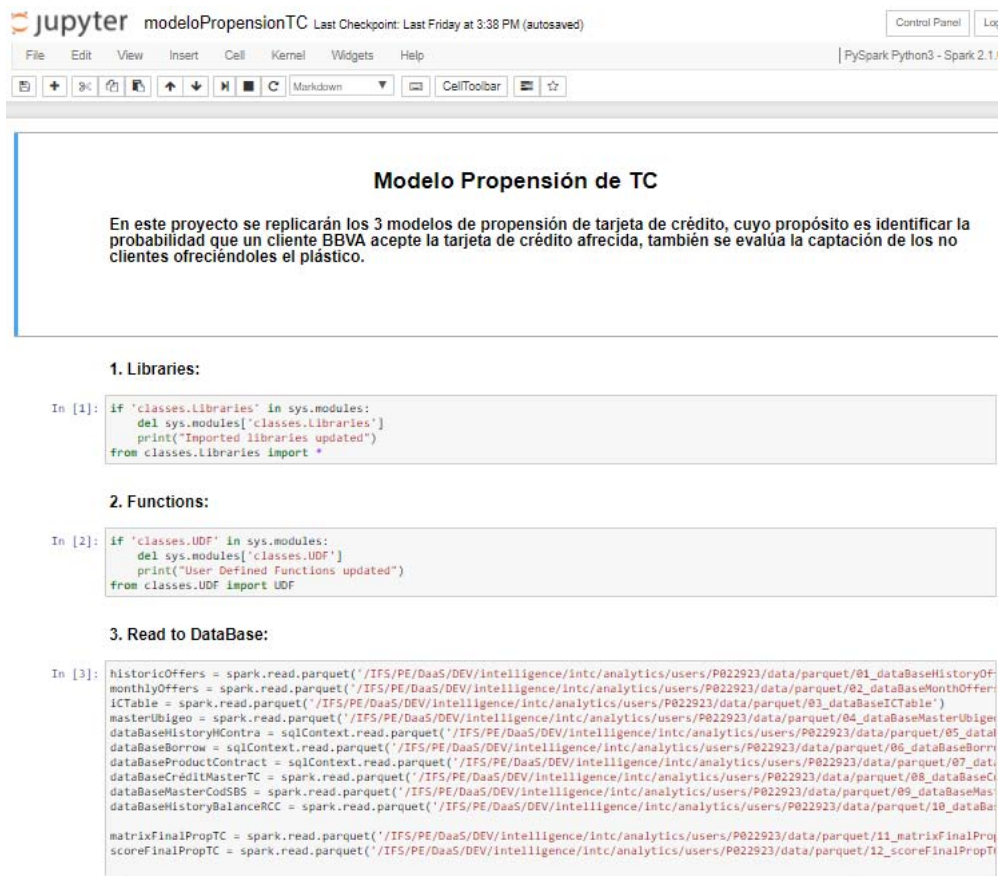


Figure 7: Notebook with the Credit Card Propensity Model

While we are working with the Scrum methodological framework, we have iterative deliveries, so each notebook which the Data Scientists release are productivized by the developers.

The Productivization implies new developments very similar to ingests to the Data Lake, with the difference which have business logic which means the business logic is in each notebook of the different Data Scientists.

Finally, finishing all the developments put in Live and meshed with the Control M programmer, the automatic process of Credit Card Campaigns is launched, the result from all the processes are flat files with the identifiers of the potential clients for acquiring a card credit, this list of customers has been sending to the different distribution channels from the financial institution (Web, ATM, Email, Social Networks).

e) Pilot Validation

In the validation from the pilot there are 2 stages:

1. *Stage in Sandbox:* The different users proceed to balance the results of their analytical models and procedures with the results obtained from the performing from the campaigns in a traditional way (35-day old process). An exact match is not expected since in many cases the models were

optimized with more reliable data and expanded the universe of possible clients to obtain credit cards.

Another characteristic from this stage is that validated notebooks are built, which will serve us in the second stage for matching the production processes. These notebooks contain automatic tables which are reading the results and matching them with the expected targets, as well as the business rules which are contemplated by the Data Scientist; Compared with the traditional software development projects these would be acceptance tests.

2. *White March Stage:* Started when the productivization of the processes performed by the developers was completed; here all Data Scientist notebooks have been implemented in jobs that will performed under Spark on a mesh orchestrated by Control M. In this way, the data resulting from the scheduled executions is going to monitoring with the validated notebooks from the previous stage. If there is some observation founded, the development is corrected and reprocessing is launched, allowing the data partition being validated to be overwritten.

f) *Deployment*

The satisfactory results from the pilot of credit card campaigns allow the Big Data environment to be deployed to other initiatives of the transformation plan, reason for this, committees are formed between the Project Management and Engineering area for developing the projects according to the profitability, complexity and impact will be realized in the Big Data environment.

The work from the Solution Architects becomes relevant when they design a document with the architectural solution model for the different use cases, this global guideline helps in the elaboration from the technical and functional designs and the own ingest implementations in the Data Lake. These solutions involve all the layers of a software development such as services or the front end, so these solution architects rely on data architects to suggest specific solution alternatives in the Data Lake or outside of it.

The architecture, data scientist and engineering communities are made available with the aim of sharing experience with other countries where the financial institution also operates. This communication reinforces the good practices established and the recommendations for initiatives.

g) *Reliability and Research*

Once the big data environment is performed, more projects and more users progressively consume it, which contributes to the formalization of a platform support front is crucial for the stability of a continuous work environment and the reliability of a platform (Nachiappan et al. , 2017).

This support team was initially centralized for Latin America where the infrastructure of all the

countries is located, but soon after, it was decentralized to the different regions providing the Engineering teams autonomy to provide access to files in sandboxes, user registrations, creation of development environments for developers, and creation of sandboxes; This autonomy improved the perception from the quality of internal services by Architecture and Engineering because the scaling of capacities is better in a distributed environment like in the case of Big Data, unlike current systems where Engineering was always an obstacle, for example increased storage.

As time goes by, new pieces come to join the Big Data environment and others to replace, in the case of Nexus by JFrog Artifactory (as a repository from configuration artifacts) or new libraries for data ingestion in the Data Lake or for measuring the quality from data in the same Data Lake; For this, Arquitectura de Datos continues with its component validation work, ensuring the correct use of each piece of the Big Data environment that comes from Spain.

The productive batch will grow while the projects and initiatives will increase, for this, the reliability of Engineering area ensures the reliability from the meshes in Control M, considering direct communication with us, the architecture team and the infrastructure commissions in Mexico.

Finally, the Big Data environment is constantly monitored locally and globally through strategic KPIs, which involves resource consumption (cpus, memory and disk), user accesses, incidence analysis and monitoring of governed objects.



Figure 8: Example of Monitoring of the Environment

For obtaining this data, we support on tools for continuous monitoring of cores, users, memory, etc., of our Big Data environment which is in continuous evolution.

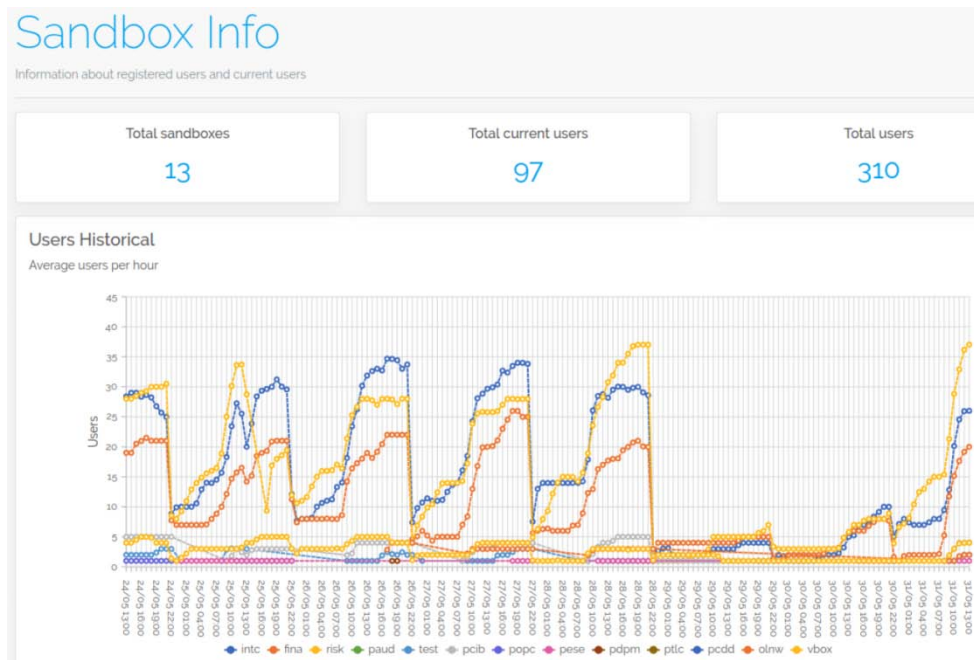
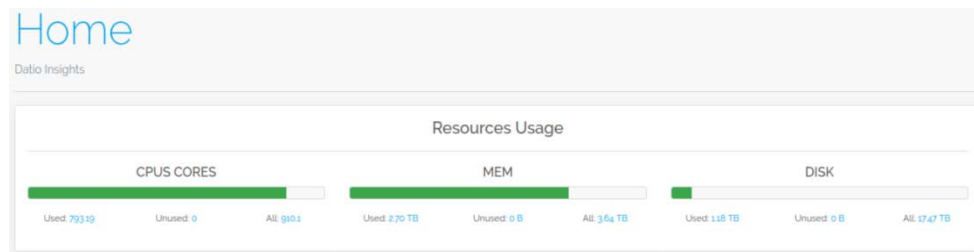


Figure 9: Tools of De Monitoring from the Consumption in Big Data Environment



The upper graph (Sandbox Info) contains the evolution from the number of active users in the different sandboxes per minute in the last 5 days; for example, at that moment we have 13 sandboxes with 97 active users from a total of 310 users which could be connect in sometime; the Y axis are the different sandboxes.

The lower graph (Home) shows in live, the cpus, memory and disk consumed in the Big Data environment; for example, the first box indicates 849 cores (cpus) from a total of 910, the second box indicates 285 TB used from a total of 364 TB, the third box indicates 1.18 TB consumed from a total of 17.47 TB.

We understand that the Platform Validation, Deployment and Reliability and Research phases described above, belong to the qualities that must be achieved in a Big Data architecture, to this are added features such as considering a distributed data processing, being scalable both in software as in hardware, being available and for distributing a large volume of data; and mainly 3 stages for the implementation of the Big Data Architecture which must be aligned to the proposals in this present article: "Obtaining data from different sources", "Real-time data

processing", "Analysis, visualization and decision making". (Quiroz et al., 2019)

III. RESULTS

As results from the integration, there are several projects working on this Big Data environment, which are productive with governed data, defined flows and with better response times in processes which were previously developed in traditional databases. With this, the business areas have more efficient products and services.

This integration of the Big data environment is the starting point for shutting down the local systems / databases, which are a problem for the organization by causing dispersed and non-consensual information between the business areas which used these data. This is solved with centralized and governed data self-service sandboxes.

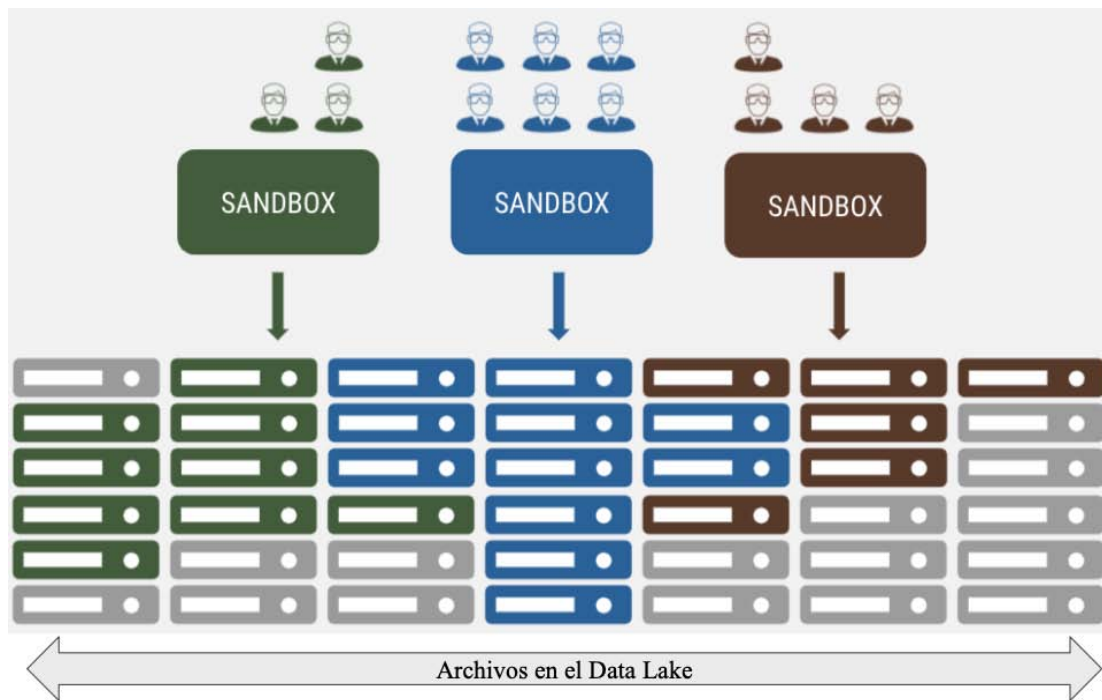


Figure 10: Design from the Distribution of Sandboxes

The first pilot from the credit card campaign was not successful, so the respective modifications were made for the following month because this campaign process is monthly. With the second attempt

to execute the campaign process in the Big Data environment, the objective was achieved, reducing from 35 days to 5 days with reliable sources, adequate, efficient and governed storage.

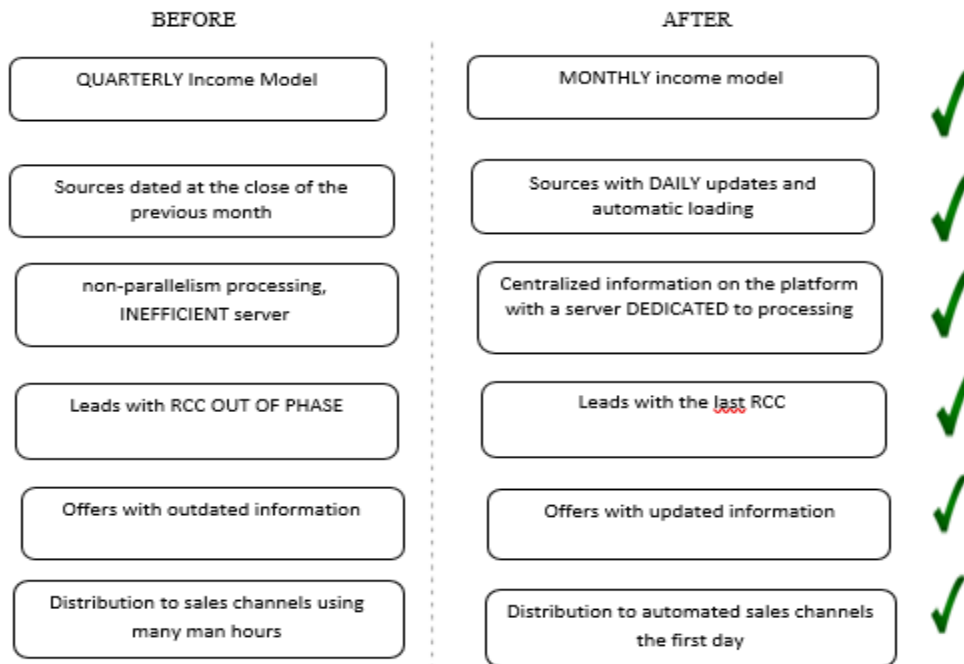


Figure 11: Qualitative Variables for Validation of the Integration

a) *Distribution to Automated Sales Channels at First Day*

The use from the agile methodological framework served a lot for the iterative deliveries of the

integration from the Big Data environment, but the present work could be the first stage for a new methodology which involves concepts from the software development, service levels and embedded pilots.

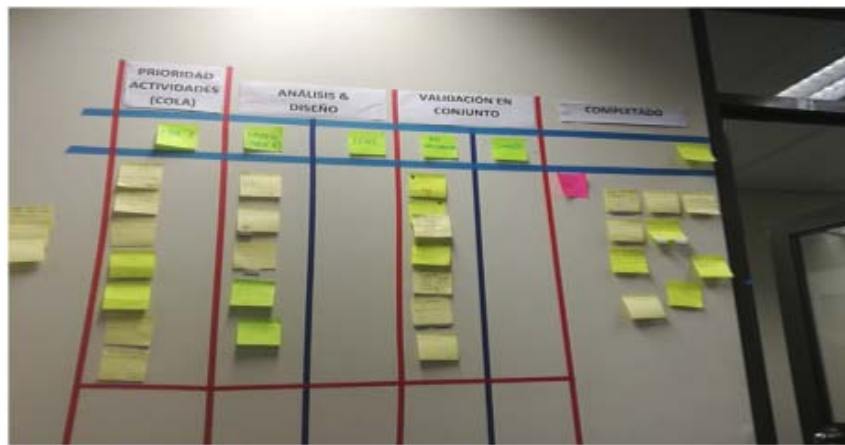


Figure 12: Scrum from the Pilot in the Integration from the Big Data

As a result, with regard to the business areas, the user is provided with data with a wide possibility of exploitation, complying with the 7 vs. The BBVA Api Market portal (BBVA et al., 2020) also considers Big Data characteristics: Volume, Speed, Variety, Value, Truthfulness, Variability, Visualization.

IV. DISCUSSION

It is important to shut down isolated systems and databases because to the overflow related to the infrastructure, maintainability and the mismanagement of the data causing the information cannot be truthful and little by little to lack value.

With the optimized campaign process, by deploying it in Big Data, the financial entity becomes competitive with other entities in the market for customers.

The Data Scientists were not 100% dedicated to the development of the pilot, this complicated many meetings for defining the rules, which in the long term, caused rework; This occurs because the respective business management areas did not want to neglect their daily operations by giving up their best resources.

Although the Data Scientists were not dedicated to the pilot, the commitment and motivation was always high and it was reflected in the rapport of the team which is vital for the success of any project.

Having a Big Data environment and a governed Data Lake provide the scenery for other types of non-financial analysis in these banking entities, but also for analysis of corporate social responsibility and / or social networks. (López et al., 2018)

V. CONCLUSIONS

The integration of the Big Data environment in the financial management contributed to the efficient generation of products and services for the Business Development area which also optimizing its processes, internal services and decision-making at the management level.

With the Big Data environment, the business areas will develop advanced analytics for visualizing and predicting the offers from new products, services, customer segmentation and behavior for credit campaigns. Inducing the business areas to be independent from the Engineering area (which was previously responsible for all development), because they will have their own Sandboxes for their own developments and a governed Data Lake.

The governed Data Lake provides us too much volume for making decisions, in addition the information is truthful by the implanted data government which has the consensus from different business areas due to the value and variety which predominate in the Data Lake.

Finally, distributed processing and storage allow the speed from access to information and generation from reports in line is according with the competition, when users have the capabilities of Business Intelligence and Big Data.

The architecture is very flexible because to its hybrid nature (which contains open code components in its structure) adopting to the new market trends, for example, there are agreements with Google to integrate components such as: Kubernetes; or with Amazon Web Service to use your endpoints and improve the access to the services from our Big Data environment. All these components allow us to expand our knowledge and enrich the role of Data Architecture.

The use of Spark allows optimizing the processing times, using traditional data engines because the times were very high. One of the campaign model processes took 24 hours to generate results, while with spark is only 60 minutes.

REFERENCES RÉFÉRENCES REFERENCIAS

1. Plase et al., 2017. A comparison of HDFS compact data formats: Avro Versus Parquet - Daiga Plase, Laila Niedrite, Romans Taranovs <https://journals.vgtu.lt/index.php/MLA/article/view/500/357>

2. BBVA et al., 2020. Api Market BBVA <https://www.bbvaapimarket.com/es/mundo-api/las-siete-v-del-big-data/>
3. Galeano et al., 2019. Las nuevas Oportunidad del Big Data para las instituciones financieras - Pedro Galeno, Daniel Peña https://www.funcas.es/wp-content/uploads/Migracion/Articulos/FUNCAS_PEE/162art07.pdf
4. Quiroz et al., 2019. Proceso de diseño de una arquitectura Big Data para el análisis de grandes volúmenes de datos e información. Miguel Quiroz Martínez, Ricardo Andrés Aguilar Duarte, Diego B. Intriago Cedeño. Revista Opuntia Brava - Universidad de Las Tunas. Cuba (2019) <http://opuntiabrava.ult.edu.cu/index.php/opuntiabrava/article/download/968/1139>
5. Nachiappan et al., 2017. Cloud storage reliability for Big Data applications: A state of the art survey - Rekha Nachiappan, Bahman Javadi, Rodrigo N. Calheiros, Kenan M. Matawie <https://www.sciencedirect.com/science/article/abs/pii/S1084804517302734>
6. López et al., 2018. Stakeholder Perceptions and Word-of-Mouth on CSR Dynamics: A Big Data Analysis from Twitter. Andrée Marie López-Fernández, Zamira Burgos Silva. <https://www.igi-global.com/article/stakeholder-perceptions-and-word-of-mouth-on-csr-dynamics/193573>



GLOBAL JOURNALS GUIDELINES HANDBOOK 2022

WWW.GLOBALJOURNALS.ORG

MEMBERSHIPS

FELLOWS/ASSOCIATES OF COMPUTER SCIENCE RESEARCH COUNCIL FCSRC/ACSRC MEMBERSHIPS

INTRODUCTION



FCSRC/ACSRC is the most prestigious membership of Global Journals accredited by Open Association of Research Society, U.S.A (OARS). The credentials of Fellow and Associate designations signify that the researcher has gained the knowledge of the fundamental and high-level concepts, and is a subject matter expert, proficient in an expertise course covering the professional code of conduct, and follows recognized standards of practice. The credentials are designated only to the researchers, scientists, and professionals that have been selected by a rigorous process by our Editorial Board and Management Board.

Associates of FCSRC/ACSRC are scientists and researchers from around the world are working on projects/researches that have huge potentials. Members support Global Journals' mission to advance technology for humanity and the profession.

FCSRC

FELLOW OF COMPUTER SCIENCE RESEARCH COUNCIL

FELLOW OF COMPUTER SCIENCE RESEARCH COUNCIL is the most prestigious membership of Global Journals. It is an award and membership granted to individuals that the Open Association of Research Society judges to have made a 'substantial contribution to the improvement of computer science, technology, and electronics engineering.

The primary objective is to recognize the leaders in research and scientific fields of the current era with a global perspective and to create a channel between them and other researchers for better exposure and knowledge sharing. Members are most eminent scientists, engineers, and technologists from all across the world. Fellows are elected for life through a peer review process on the basis of excellence in the respective domain. There is no limit on the number of new nominations made in any year. Each year, the Open Association of Research Society elect up to 12 new Fellow Members.



BENEFIT

TO THE INSTITUTION

GET LETTER OF APPRECIATION

Global Journals sends a letter of appreciation of author to the Dean or CEO of the University or Company of which author is a part, signed by editor in chief or chief author.



EXCLUSIVE NETWORK

GET ACCESS TO A CLOSED NETWORK

A FCSRC member gets access to a closed network of Tier 1 researchers and scientists with direct communication channel through our website. Fellows can reach out to other members or researchers directly. They should also be open to reaching out by other.

Career

Credibility

Exclusive

Reputation



CERTIFICATE

CERTIFICATE, LOR AND LASER-MOMENTO

Fellows receive a printed copy of a certificate signed by our Chief Author that may be used for academic purposes and a personal recommendation letter to the dean of member's university.

Career

Credibility

Exclusive

Reputation



DESIGNATION

GET HONORED TITLE OF MEMBERSHIP

Fellows can use the honored title of membership. The "FCSRC" is an honored title which is accorded to a person's name viz. Dr. John E. Hall, Ph.D., FCSRC or William Walldroff, M.S., FCSRC.

Career

Credibility

Exclusive

Reputation

RECOGNITION ON THE PLATFORM

BETTER VISIBILITY AND CITATION

All the Fellow members of FCSRC get a badge of "Leading Member of Global Journals" on the Research Community that distinguishes them from others. Additionally, the profile is also partially maintained by our team for better visibility and citation. All fellows get a dedicated page on the website with their biography.

Career

Credibility

Reputation

FUTURE WORK

GET DISCOUNTS ON THE FUTURE PUBLICATIONS

Fellows receive discounts on future publications with Global Journals up to 60%. Through our recommendation programs, members also receive discounts on publications made with OARS affiliated organizations.

Career

Financial



GJ ACCOUNT

UNLIMITED FORWARD OF EMAILS

Fellows get secure and fast GJ work emails with unlimited forward of emails that they may use them as their primary email. For example, john [AT] globaljournals [DOT] org.

Career

Credibility

Reputation



PREMIUM TOOLS

ACCESS TO ALL THE PREMIUM TOOLS

To take future researches to the zenith, fellows receive access to all the premium tools that Global Journals have to offer along with the partnership with some of the best marketing leading tools out there.

Financial

CONFERENCES & EVENTS

ORGANIZE SEMINAR/CONFERENCE

Fellows are authorized to organize symposium/seminar/conference on behalf of Global Journal Incorporation (USA). They can also participate in the same organized by another institution as representative of Global Journal. In both the cases, it is mandatory for him to discuss with us and obtain our consent. Additionally, they get free research conferences (and others) alerts.

Career

Credibility

Financial

EARLY INVITATIONS

EARLY INVITATIONS TO ALL THE SYMPOSIUMS, SEMINARS, CONFERENCES

All fellows receive the early invitations to all the symposiums, seminars, conferences and webinars hosted by Global Journals in their subject.

Exclusive



PUBLISHING ARTICLES & BOOKS

EARN 60% OF SALES PROCEEDS

Fellows can publish articles (limited) without any fees. Also, they can earn up to 70% of sales proceeds from the sale of reference/review books/literature/publishing of research paper. The FCSRC member can decide its price and we can help in making the right decision.

Exclusive

Financial

REVIEWERS

GET A REMUNERATION OF 15% OF AUTHOR FEES

Fellow members are eligible to join as a paid peer reviewer at Global Journals Incorporation (USA) and can get a remuneration of 15% of author fees, taken from the author of a respective paper.

Financial

ACCESS TO EDITORIAL BOARD

BECOME A MEMBER OF THE EDITORIAL BOARD

Fellows may join as a member of the Editorial Board of Global Journals Incorporation (USA) after successful completion of three years as Fellow and as Peer Reviewer. Additionally, Fellows get a chance to nominate other members for Editorial Board.

Career

Credibility

Exclusive

Reputation

AND MUCH MORE

GET ACCESS TO SCIENTIFIC MUSEUMS AND OBSERVATORIES ACROSS THE GLOBE

All members get access to 5 selected scientific museums and observatories across the globe. All researches published with Global Journals will be kept under deep archival facilities across regions for future protections and disaster recovery. They get 10 GB free secure cloud access for storing research files.

ASSOCIATE OF COMPUTER SCIENCE RESEARCH COUNCIL

ASSOCIATE OF COMPUTER SCIENCE RESEARCH COUNCIL is the membership of Global Journals awarded to individuals that the Open Association of Research Society judges to have made a 'substantial contribution to the improvement of computer science, technology, and electronics engineering.

The primary objective is to recognize the leaders in research and scientific fields of the current era with a global perspective and to create a channel between them and other researchers for better exposure and knowledge sharing. Members are most eminent scientists, engineers, and technologists from all across the world. Associate membership can later be promoted to Fellow Membership. Associates are elected for life through a peer review process on the basis of excellence in the respective domain. There is no limit on the number of new nominations made in any year. Each year, the Open Association of Research Society elect up to 12 new Associate Members.



BENEFIT

TO THE INSTITUTION

GET LETTER OF APPRECIATION

Global Journals sends a letter of appreciation of author to the Dean or CEO of the University or Company of which author is a part, signed by editor in chief or chief author.



EXCLUSIVE NETWORK

GET ACCESS TO A CLOSED NETWORK

A ACSRC member gets access to a closed network of Tier 2 researchers and scientists with direct communication channel through our website. Associates can reach out to other members or researchers directly. They should also be open to reaching out by other.

Career

Credibility

Exclusive

Reputation



CERTIFICATE

CERTIFICATE, LOR AND LASER-MOMENTO

Associates receive a printed copy of a certificate signed by our Chief Author that may be used for academic purposes and a personal recommendation letter to the dean of member's university.

Career

Credibility

Exclusive

Reputation



DESIGNATION

GET HONORED TITLE OF MEMBERSHIP

Associates can use the honored title of membership. The "ACSRC" is an honored title which is accorded to a person's name viz. Dr. John E. Hall, Ph.D., ACSRC or William Walldroff, M.S., ACSRC.

Career

Credibility

Exclusive

Reputation

RECOGNITION ON THE PLATFORM

BETTER VISIBILITY AND CITATION

All the Associate members of ACSRC get a badge of "Leading Member of Global Journals" on the Research Community that distinguishes them from others. Additionally, the profile is also partially maintained by our team for better visibility and citation.

Career

Credibility

Reputation

FUTURE WORK

GET DISCOUNTS ON THE FUTURE PUBLICATIONS

Associates receive discounts on future publications with Global Journals up to 30%. Through our recommendation programs, members also receive discounts on publications made with OARS affiliated organizations.

Career

Financial



GJ ACCOUNT

UNLIMITED FORWARD OF EMAILS

Associates get secure and fast GJ work emails with 5GB forward of emails that they may use them as their primary email. For example, john [AT] globaljournals [DOT] org.

Career

Credibility

Reputation



PREMIUM TOOLS

ACCESS TO ALL THE PREMIUM TOOLS

To take future researches to the zenith, associates receive access to all the premium tools that Global Journals have to offer along with the partnership with some of the best marketing leading tools out there.

Financial

CONFERENCES & EVENTS

ORGANIZE SEMINAR/CONFERENCE

Associates are authorized to organize symposium/seminar/conference on behalf of Global Journal Incorporation (USA). They can also participate in the same organized by another institution as representative of Global Journal. In both the cases, it is mandatory for him to discuss with us and obtain our consent. Additionally, they get free research conferences (and others) alerts.

Career

Credibility

Financial

EARLY INVITATIONS

EARLY INVITATIONS TO ALL THE SYMPOSIUMS, SEMINARS, CONFERENCES

All associates receive the early invitations to all the symposiums, seminars, conferences and webinars hosted by Global Journals in their subject.

Exclusive





PUBLISHING ARTICLES & BOOKS

EARN 30-40% OF SALES PROCEEDS

Associates can publish articles (limited) without any fees. Also, they can earn up to 30-40% of sales proceeds from the sale of reference/review books/literature/publishing of research paper.

Exclusive

Financial

REVIEWERS

GET A REMUNERATION OF 15% OF AUTHOR FEES

Associate members are eligible to join as a paid peer reviewer at Global Journals Incorporation (USA) and can get a remuneration of 15% of author fees, taken from the author of a respective paper.

Financial

AND MUCH MORE

GET ACCESS TO SCIENTIFIC MUSEUMS AND OBSERVATORIES ACROSS THE GLOBE

All members get access to 2 selected scientific museums and observatories across the globe. All researches published with Global Journals will be kept under deep archival facilities across regions for future protections and disaster recovery. They get 5 GB free secure cloud access for storing research files.



ASSOCIATE	FELLOW	RESEARCH GROUP	BASIC
<p>\$4800 lifetime designation</p> <hr/> <p>Certificate, LoR and Momento 2 discounted publishing/year Gradation of Research 10 research contacts/day 1 GB Cloud Storage GJ Community Access</p>	<p>\$6800 lifetime designation</p> <hr/> <p>Certificate, LoR and Momento Unlimited discounted publishing/year Gradation of Research Unlimited research contacts/day 5 GB Cloud Storage Online Presense Assistance GJ Community Access</p>	<p>\$12500.00 organizational</p> <hr/> <p>Certificates, LoRs and Momentos Unlimited free publishing/year Gradation of Research Unlimited research contacts/day Unlimited Cloud Storage Online Presense Assistance GJ Community Access</p>	<p>APC per article</p> <hr/> <p>GJ Community Access</p>



PREFERRED AUTHOR GUIDELINES

We accept the manuscript submissions in any standard (generic) format.

We typeset manuscripts using advanced typesetting tools like Adobe In Design, CorelDraw, TeXnicCenter, and TeXStudio. We usually recommend authors submit their research using any standard format they are comfortable with, and let Global Journals do the rest.

Alternatively, you can download our basic template from <https://globaljournals.org/Template.zip>

Authors should submit their complete paper/article, including text illustrations, graphics, conclusions, artwork, and tables. Authors who are not able to submit manuscript using the form above can email the manuscript department at submit@globaljournals.org or get in touch with chiefeditor@globaljournals.org if they wish to send the abstract before submission.

BEFORE AND DURING SUBMISSION

Authors must ensure the information provided during the submission of a paper is authentic. Please go through the following checklist before submitting:

1. Authors must go through the complete author guideline and understand and *agree to Global Journals' ethics and code of conduct*, along with author responsibilities.
2. Authors must accept the privacy policy, terms, and conditions of Global Journals.
3. Ensure corresponding author's email address and postal address are accurate and reachable.
4. Manuscript to be submitted must include keywords, an abstract, a paper title, co-author(s) names and details (email address, name, phone number, and institution), figures and illustrations in vector format including appropriate captions, tables, including titles and footnotes, a conclusion, results, acknowledgments and references.
5. Authors should submit paper in a ZIP archive if any supplementary files are required along with the paper.
6. Proper permissions must be acquired for the use of any copyrighted material.
7. Manuscript submitted *must not have been submitted or published elsewhere* and all authors must be aware of the submission.

Declaration of Conflicts of Interest

It is required for authors to declare all financial, institutional, and personal relationships with other individuals and organizations that could influence (bias) their research.

POLICY ON PLAGIARISM

Plagiarism is not acceptable in Global Journals submissions at all.

Plagiarized content will not be considered for publication. We reserve the right to inform authors' institutions about plagiarism detected either before or after publication. If plagiarism is identified, we will follow COPE guidelines:

Authors are solely responsible for all the plagiarism that is found. The author must not fabricate, falsify or plagiarize existing research data. The following, if copied, will be considered plagiarism:

- Words (language)
- Ideas
- Findings
- Writings
- Diagrams
- Graphs
- Illustrations
- Lectures



- Printed material
- Graphic representations
- Computer programs
- Electronic material
- Any other original work

AUTHORSHIP POLICIES

Global Journals follows the definition of authorship set up by the Open Association of Research Society, USA. According to its guidelines, authorship criteria must be based on:

1. Substantial contributions to the conception and acquisition of data, analysis, and interpretation of findings.
2. Drafting the paper and revising it critically regarding important academic content.
3. Final approval of the version of the paper to be published.

Changes in Authorship

The corresponding author should mention the name and complete details of all co-authors during submission and in manuscript. We support addition, rearrangement, manipulation, and deletions in authors list till the early view publication of the journal. We expect that corresponding author will notify all co-authors of submission. We follow COPE guidelines for changes in authorship.

Copyright

During submission of the manuscript, the author is confirming an exclusive license agreement with Global Journals which gives Global Journals the authority to reproduce, reuse, and republish authors' research. We also believe in flexible copyright terms where copyright may remain with authors/employers/institutions as well. Contact your editor after acceptance to choose your copyright policy. You may follow this form for copyright transfers.

Appealing Decisions

Unless specified in the notification, the Editorial Board's decision on publication of the paper is final and cannot be appealed before making the major change in the manuscript.

Acknowledgments

Contributors to the research other than authors credited should be mentioned in Acknowledgments. The source of funding for the research can be included. Suppliers of resources may be mentioned along with their addresses.

Declaration of funding sources

Global Journals is in partnership with various universities, laboratories, and other institutions worldwide in the research domain. Authors are requested to disclose their source of funding during every stage of their research, such as making analysis, performing laboratory operations, computing data, and using institutional resources, from writing an article to its submission. This will also help authors to get reimbursements by requesting an open access publication letter from Global Journals and submitting to the respective funding source.

PREPARING YOUR MANUSCRIPT

Authors can submit papers and articles in an acceptable file format: MS Word (doc, docx), LaTeX (.tex, .zip or .rar including all of your files), Adobe PDF (.pdf), rich text format (.rtf), simple text document (.txt), Open Document Text (.odt), and Apple Pages (.pages). Our professional layout editors will format the entire paper according to our official guidelines. This is one of the highlights of publishing with Global Journals—authors should not be concerned about the formatting of their paper. Global Journals accepts articles and manuscripts in every major language, be it Spanish, Chinese, Japanese, Portuguese, Russian, French, German, Dutch, Italian, Greek, or any other national language, but the title, subtitle, and abstract should be in English. This will facilitate indexing and the pre-peer review process.

The following is the official style and template developed for publication of a research paper. Authors are not required to follow this style during the submission of the paper. It is just for reference purposes.



Manuscript Style Instruction (Optional)

- Microsoft Word Document Setting Instructions.
- Font type of all text should be Swis721 Lt BT.
- Page size: 8.27" x 11", left margin: 0.65, right margin: 0.65, bottom margin: 0.75.
- Paper title should be in one column of font size 24.
- Author name in font size of 11 in one column.
- Abstract: font size 9 with the word "Abstract" in bold italics.
- Main text: font size 10 with two justified columns.
- Two columns with equal column width of 3.38 and spacing of 0.2.
- First character must be three lines drop-capped.
- The paragraph before spacing of 1 pt and after of 0 pt.
- Line spacing of 1 pt.
- Large images must be in one column.
- The names of first main headings (Heading 1) must be in Roman font, capital letters, and font size of 10.
- The names of second main headings (Heading 2) must not include numbers and must be in italics with a font size of 10.

Structure and Format of Manuscript

The recommended size of an original research paper is under 15,000 words and review papers under 7,000 words. Research articles should be less than 10,000 words. Research papers are usually longer than review papers. Review papers are reports of significant research (typically less than 7,000 words, including tables, figures, and references)

A research paper must include:

- a) A title which should be relevant to the theme of the paper.
- b) A summary, known as an abstract (less than 150 words), containing the major results and conclusions.
- c) Up to 10 keywords that precisely identify the paper's subject, purpose, and focus.
- d) An introduction, giving fundamental background objectives.
- e) Resources and techniques with sufficient complete experimental details (wherever possible by reference) to permit repetition, sources of information must be given, and numerical methods must be specified by reference.
- f) Results which should be presented concisely by well-designed tables and figures.
- g) Suitable statistical data should also be given.
- h) All data must have been gathered with attention to numerical detail in the planning stage.

Design has been recognized to be essential to experiments for a considerable time, and the editor has decided that any paper that appears not to have adequate numerical treatments of the data will be returned unrefereed.

- i) Discussion should cover implications and consequences and not just recapitulate the results; conclusions should also be summarized.
- j) There should be brief acknowledgments.
- k) There ought to be references in the conventional format. Global Journals recommends APA format.

Authors should carefully consider the preparation of papers to ensure that they communicate effectively. Papers are much more likely to be accepted if they are carefully designed and laid out, contain few or no errors, are summarizing, and follow instructions. They will also be published with much fewer delays than those that require much technical and editorial correction.

The Editorial Board reserves the right to make literary corrections and suggestions to improve brevity.



FORMAT STRUCTURE

It is necessary that authors take care in submitting a manuscript that is written in simple language and adheres to published guidelines.

All manuscripts submitted to Global Journals should include:

Title

The title page must carry an informative title that reflects the content, a running title (less than 45 characters together with spaces), names of the authors and co-authors, and the place(s) where the work was carried out.

Author details

The full postal address of any related author(s) must be specified.

Abstract

The abstract is the foundation of the research paper. It should be clear and concise and must contain the objective of the paper and inferences drawn. It is advised to not include big mathematical equations or complicated jargon.

Many researchers searching for information online will use search engines such as Google, Yahoo or others. By optimizing your paper for search engines, you will amplify the chance of someone finding it. In turn, this will make it more likely to be viewed and cited in further works. Global Journals has compiled these guidelines to facilitate you to maximize the web-friendliness of the most public part of your paper.

Keywords

A major lynchpin of research work for the writing of research papers is the keyword search, which one will employ to find both library and internet resources. Up to eleven keywords or very brief phrases have to be given to help data retrieval, mining, and indexing.

One must be persistent and creative in using keywords. An effective keyword search requires a strategy: planning of a list of possible keywords and phrases to try.

Choice of the main keywords is the first tool of writing a research paper. Research paper writing is an art. Keyword search should be as strategic as possible.

One should start brainstorming lists of potential keywords before even beginning searching. Think about the most important concepts related to research work. Ask, "What words would a source have to include to be truly valuable in a research paper?" Then consider synonyms for the important words.

It may take the discovery of only one important paper to steer in the right keyword direction because, in most databases, the keywords under which a research paper is abstracted are listed with the paper.

Numerical Methods

Numerical methods used should be transparent and, where appropriate, supported by references.

Abbreviations

Authors must list all the abbreviations used in the paper at the end of the paper or in a separate table before using them.

Formulas and equations

Authors are advised to submit any mathematical equation using either MathJax, KaTeX, or LaTeX, or in a very high-quality image.

Tables, Figures, and Figure Legends

Tables: Tables should be cautiously designed, uncrowned, and include only essential data. Each must have an Arabic number, e.g., Table 4, a self-explanatory caption, and be on a separate sheet. Authors must submit tables in an editable format and not as images. References to these tables (if any) must be mentioned accurately.



Figures

Figures are supposed to be submitted as separate files. Always include a citation in the text for each figure using Arabic numbers, e.g., Fig. 4. Artwork must be submitted online in vector electronic form or by emailing it.

PREPARATION OF ELETRONIC FIGURES FOR PUBLICATION

Although low-quality images are sufficient for review purposes, print publication requires high-quality images to prevent the final product being blurred or fuzzy. Submit (possibly by e-mail) EPS (line art) or TIFF (halftone/ photographs) files only. MS PowerPoint and Word Graphics are unsuitable for printed pictures. Avoid using pixel-oriented software. Scans (TIFF only) should have a resolution of at least 350 dpi (halftone) or 700 to 1100 dpi (line drawings). Please give the data for figures in black and white or submit a Color Work Agreement form. EPS files must be saved with fonts embedded (and with a TIFF preview, if possible).

For scanned images, the scanning resolution at final image size ought to be as follows to ensure good reproduction: line art: >650 dpi; halftones (including gel photographs): >350 dpi; figures containing both halftone and line images: >650 dpi.

Color charges: Authors are advised to pay the full cost for the reproduction of their color artwork. Hence, please note that if there is color artwork in your manuscript when it is accepted for publication, we would require you to complete and return a Color Work Agreement form before your paper can be published. Also, you can email your editor to remove the color fee after acceptance of the paper.

TIPS FOR WRITING A GOOD QUALITY COMPUTER SCIENCE RESEARCH PAPER

Techniques for writing a good quality computer science research paper:

1. Choosing the topic: In most cases, the topic is selected by the interests of the author, but it can also be suggested by the guides. You can have several topics, and then judge which you are most comfortable with. This may be done by asking several questions of yourself, like "Will I be able to carry out a search in this area? Will I find all necessary resources to accomplish the search? Will I be able to find all information in this field area?" If the answer to this type of question is "yes," then you ought to choose that topic. In most cases, you may have to conduct surveys and visit several places. Also, you might have to do a lot of work to find all the rises and falls of the various data on that subject. Sometimes, detailed information plays a vital role, instead of short information. Evaluators are human: The first thing to remember is that evaluators are also human beings. They are not only meant for rejecting a paper. They are here to evaluate your paper. So present your best aspect.

2. Think like evaluators: If you are in confusion or getting demotivated because your paper may not be accepted by the evaluators, then think, and try to evaluate your paper like an evaluator. Try to understand what an evaluator wants in your research paper, and you will automatically have your answer. Make blueprints of paper: The outline is the plan or framework that will help you to arrange your thoughts. It will make your paper logical. But remember that all points of your outline must be related to the topic you have chosen.

3. Ask your guides: If you are having any difficulty with your research, then do not hesitate to share your difficulty with your guide (if you have one). They will surely help you out and resolve your doubts. If you can't clarify what exactly you require for your work, then ask your supervisor to help you with an alternative. He or she might also provide you with a list of essential readings.

4. Use of computer is recommended: As you are doing research in the field of computer science then this point is quite obvious. Use right software: Always use good quality software packages. If you are not capable of judging good software, then you can lose the quality of your paper unknowingly. There are various programs available to help you which you can get through the internet.

5. Use the internet for help: An excellent start for your paper is using Google. It is a wondrous search engine, where you can have your doubts resolved. You may also read some answers for the frequent question of how to write your research paper or find a model research paper. You can download books from the internet. If you have all the required books, place importance on reading, selecting, and analyzing the specified information. Then sketch out your research paper. Use big pictures: You may use encyclopedias like Wikipedia to get pictures with the best resolution. At Global Journals, you should strictly follow here.



6. Bookmarks are useful: When you read any book or magazine, you generally use bookmarks, right? It is a good habit which helps to not lose your continuity. You should always use bookmarks while searching on the internet also, which will make your search easier.

7. Revise what you wrote: When you write anything, always read it, summarize it, and then finalize it.

8. Make every effort: Make every effort to mention what you are going to write in your paper. That means always have a good start. Try to mention everything in the introduction—what is the need for a particular research paper. Polish your work with good writing skills and always give an evaluator what he wants. Make backups: When you are going to do any important thing like making a research paper, you should always have backup copies of it either on your computer or on paper. This protects you from losing any portion of your important data.

9. Produce good diagrams of your own: Always try to include good charts or diagrams in your paper to improve quality. Using several unnecessary diagrams will degrade the quality of your paper by creating a hodgepodge. So always try to include diagrams which were made by you to improve the readability of your paper. Use of direct quotes: When you do research relevant to literature, history, or current affairs, then use of quotes becomes essential, but if the study is relevant to science, use of quotes is not preferable.

10. Use proper verb tense: Use proper verb tenses in your paper. Use past tense to present those events that have happened. Use present tense to indicate events that are going on. Use future tense to indicate events that will happen in the future. Use of wrong tenses will confuse the evaluator. Avoid sentences that are incomplete.

11. Pick a good study spot: Always try to pick a spot for your research which is quiet. Not every spot is good for studying.

12. Know what you know: Always try to know what you know by making objectives, otherwise you will be confused and unable to achieve your target.

13. Use good grammar: Always use good grammar and words that will have a positive impact on the evaluator; use of good vocabulary does not mean using tough words which the evaluator has to find in a dictionary. Do not fragment sentences. Eliminate one-word sentences. Do not ever use a big word when a smaller one would suffice.

Verbs have to be in agreement with their subjects. In a research paper, do not start sentences with conjunctions or finish them with prepositions. When writing formally, it is advisable to never split an infinitive because someone will (wrongly) complain. Avoid clichés like a disease. Always shun irritating alliteration. Use language which is simple and straightforward. Put together a neat summary.

14. Arrangement of information: Each section of the main body should start with an opening sentence, and there should be a changeover at the end of the section. Give only valid and powerful arguments for your topic. You may also maintain your arguments with records.

15. Never start at the last minute: Always allow enough time for research work. Leaving everything to the last minute will degrade your paper and spoil your work.

16. Multitasking in research is not good: Doing several things at the same time is a bad habit in the case of research activity. Research is an area where everything has a particular time slot. Divide your research work into parts, and do a particular part in a particular time slot.

17. Never copy others' work: Never copy others' work and give it your name because if the evaluator has seen it anywhere, you will be in trouble. Take proper rest and food: No matter how many hours you spend on your research activity, if you are not taking care of your health, then all your efforts will have been in vain. For quality research, take proper rest and food.

18. Go to seminars: Attend seminars if the topic is relevant to your research area. Utilize all your resources.

19. Refresh your mind after intervals: Try to give your mind a rest by listening to soft music or sleeping in intervals. This will also improve your memory. Acquire colleagues: Always try to acquire colleagues. No matter how sharp you are, if you acquire colleagues, they can give you ideas which will be helpful to your research.



20. Think technically: Always think technically. If anything happens, search for its reasons, benefits, and demerits. Think and then print: When you go to print your paper, check that tables are not split, headings are not detached from their descriptions, and page sequence is maintained.

21. Adding unnecessary information: Do not add unnecessary information like "I have used MS Excel to draw graphs." Irrelevant and inappropriate material is superfluous. Foreign terminology and phrases are not apropos. One should never take a broad view. Analogy is like feathers on a snake. Use words properly, regardless of how others use them. Remove quotations. Puns are for kids, not grunt readers. Never oversimplify: When adding material to your research paper, never go for oversimplification; this will definitely irritate the evaluator. Be specific. Never use rhythmic redundancies. Contractions shouldn't be used in a research paper. Comparisons are as terrible as clichés. Give up ampersands, abbreviations, and so on. Remove commas that are not necessary. Parenthetical words should be between brackets or commas. Understatement is always the best way to put forward earth-shaking thoughts. Give a detailed literary review.

22. Report concluded results: Use concluded results. From raw data, filter the results, and then conclude your studies based on measurements and observations taken. An appropriate number of decimal places should be used. Parenthetical remarks are prohibited here. Proofread carefully at the final stage. At the end, give an outline to your arguments. Spot perspectives of further study of the subject. Justify your conclusion at the bottom sufficiently, which will probably include examples.

23. Upon conclusion: Once you have concluded your research, the next most important step is to present your findings. Presentation is extremely important as it is the definite medium through which your research is going to be in print for the rest of the crowd. Care should be taken to categorize your thoughts well and present them in a logical and neat manner. A good quality research paper format is essential because it serves to highlight your research paper and bring to light all necessary aspects of your research.

INFORMAL GUIDELINES OF RESEARCH PAPER WRITING

Key points to remember:

- Submit all work in its final form.
- Write your paper in the form which is presented in the guidelines using the template.
- Please note the criteria peer reviewers will use for grading the final paper.

Final points:

One purpose of organizing a research paper is to let people interpret your efforts selectively. The journal requires the following sections, submitted in the order listed, with each section starting on a new page:

The introduction: This will be compiled from reference matter and reflect the design processes or outline of basis that directed you to make a study. As you carry out the process of study, the method and process section will be constructed like that. The results segment will show related statistics in nearly sequential order and direct reviewers to similar intellectual paths throughout the data that you gathered to carry out your study.

The discussion section:

This will provide understanding of the data and projections as to the implications of the results. The use of good quality references throughout the paper will give the effort trustworthiness by representing an alertness to prior workings.

Writing a research paper is not an easy job, no matter how trouble-free the actual research or concept. Practice, excellent preparation, and controlled record-keeping are the only means to make straightforward progression.

General style:

Specific editorial column necessities for compliance of a manuscript will always take over from directions in these general guidelines.

To make a paper clear: Adhere to recommended page limits.



Mistakes to avoid:

- Insertion of a title at the foot of a page with subsequent text on the next page.
- Separating a table, chart, or figure—confine each to a single page.
- Submitting a manuscript with pages out of sequence.
- In every section of your document, use standard writing style, including articles ("a" and "the").
- Keep paying attention to the topic of the paper.
- Use paragraphs to split each significant point (excluding the abstract).
- Align the primary line of each section.
- Present your points in sound order.
- Use present tense to report well-accepted matters.
- Use past tense to describe specific results.
- Do not use familiar wording; don't address the reviewer directly. Don't use slang or superlatives.
- Avoid use of extra pictures—include only those figures essential to presenting results.

Title page:

Choose a revealing title. It should be short and include the name(s) and address(es) of all authors. It should not have acronyms or abbreviations or exceed two printed lines.

Abstract: This summary should be two hundred words or less. It should clearly and briefly explain the key findings reported in the manuscript and must have precise statistics. It should not have acronyms or abbreviations. It should be logical in itself. Do not cite references at this point.

An abstract is a brief, distinct paragraph summary of finished work or work in development. In a minute or less, a reviewer can be taught the foundation behind the study, common approaches to the problem, relevant results, and significant conclusions or new questions.

Write your summary when your paper is completed because how can you write the summary of anything which is not yet written? Wealth of terminology is very essential in abstract. Use comprehensive sentences, and do not sacrifice readability for brevity; you can maintain it succinctly by phrasing sentences so that they provide more than a lone rationale. The author can at this moment go straight to shortening the outcome. Sum up the study with the subsequent elements in any summary. Try to limit the initial two items to no more than one line each.

Reason for writing the article—theory, overall issue, purpose.

- Fundamental goal.
- To-the-point depiction of the research.
- Consequences, including definite statistics—if the consequences are quantitative in nature, account for this; results of any numerical analysis should be reported. Significant conclusions or questions that emerge from the research.

Approach:

- Single section and succinct.
- An outline of the job done is always written in past tense.
- Concentrate on shortening results—limit background information to a verdict or two.
- Exact spelling, clarity of sentences and phrases, and appropriate reporting of quantities (proper units, important statistics) are just as significant in an abstract as they are anywhere else.

Introduction:

The introduction should "introduce" the manuscript. The reviewer should be presented with sufficient background information to be capable of comprehending and calculating the purpose of your study without having to refer to other works. The basis for the study should be offered. Give the most important references, but avoid making a comprehensive appraisal of the topic. Describe the problem visibly. If the problem is not acknowledged in a logical, reasonable way, the reviewer will give no attention to your results. Speak in common terms about techniques used to explain the problem, if needed, but do not present any particulars about the protocols here.



The following approach can create a valuable beginning:

- Explain the value (significance) of the study.
- Defend the model—why did you employ this particular system or method? What is its compensation? Remark upon its appropriateness from an abstract point of view as well as pointing out sensible reasons for using it.
- Present a justification. State your particular theory(-ies) or aim(s), and describe the logic that led you to choose them.
- Briefly explain the study's tentative purpose and how it meets the declared objectives.

Approach:

Use past tense except for when referring to recognized facts. After all, the manuscript will be submitted after the entire job is done. Sort out your thoughts; manufacture one key point for every section. If you make the four points listed above, you will need at least four paragraphs. Present surrounding information only when it is necessary to support a situation. The reviewer does not desire to read everything you know about a topic. Shape the theory specifically—do not take a broad view.

As always, give awareness to spelling, simplicity, and correctness of sentences and phrases.

Procedures (methods and materials):

This part is supposed to be the easiest to carve if you have good skills. A soundly written procedures segment allows a capable scientist to replicate your results. Present precise information about your supplies. The suppliers and clarity of reagents can be helpful bits of information. Present methods in sequential order, but linked methodologies can be grouped as a segment. Be concise when relating the protocols. Attempt to give the least amount of information that would permit another capable scientist to replicate your outcome, but be cautious that vital information is integrated. The use of subheadings is suggested and ought to be synchronized with the results section.

When a technique is used that has been well-described in another section, mention the specific item describing the way, but draw the basic principle while stating the situation. The purpose is to show all particular resources and broad procedures so that another person may use some or all of the methods in one more study or referee the scientific value of your work. It is not to be a step-by-step report of the whole thing you did, nor is a methods section a set of orders.

Materials:

Materials may be reported in part of a section or else they may be recognized along with your measures.

Methods:

- Report the method and not the particulars of each process that engaged the same methodology.
- Describe the method entirely.
- To be succinct, present methods under headings dedicated to specific dealings or groups of measures.
- Simplify—detail how procedures were completed, not how they were performed on a particular day.
- If well-known procedures were used, account for the procedure by name, possibly with a reference, and that's all.

Approach:

It is embarrassing to use vigorous voice when documenting methods without using first person, which would focus the reviewer's interest on the researcher rather than the job. As a result, when writing up the methods, most authors use third person passive voice.

Use standard style in this and every other part of the paper—avoid familiar lists, and use full sentences.

What to keep away from:

- Resources and methods are not a set of information.
- Skip all descriptive information and surroundings—save it for the argument.
- Leave out information that is immaterial to a third party.



Results:

The principle of a results segment is to present and demonstrate your conclusion. Create this part as entirely objective details of the outcome, and save all understanding for the discussion.

The page length of this segment is set by the sum and types of data to be reported. Use statistics and tables, if suitable, to present consequences most efficiently.

You must clearly differentiate material which would usually be incorporated in a study editorial from any unprocessed data or additional appendix matter that would not be available. In fact, such matters should not be submitted at all except if requested by the instructor.

Content:

- Sum up your conclusions in text and demonstrate them, if suitable, with figures and tables.
- In the manuscript, explain each of your consequences, and point the reader to remarks that are most appropriate.
- Present a background, such as by describing the question that was addressed by creation of an exacting study.
- Explain results of control experiments and give remarks that are not accessible in a prescribed figure or table, if appropriate.
- Examine your data, then prepare the analyzed (transformed) data in the form of a figure (graph), table, or manuscript.

What to stay away from:

- Do not discuss or infer your outcome, report surrounding information, or try to explain anything.
- Do not include raw data or intermediate calculations in a research manuscript.
- Do not present similar data more than once.
- A manuscript should complement any figures or tables, not duplicate information.
- Never confuse figures with tables—there is a difference.

Approach:

As always, use past tense when you submit your results, and put the whole thing in a reasonable order.

Put figures and tables, appropriately numbered, in order at the end of the report.

If you desire, you may place your figures and tables properly within the text of your results section.

Figures and tables:

If you put figures and tables at the end of some details, make certain that they are visibly distinguished from any attached appendix materials, such as raw facts. Whatever the position, each table must be titled, numbered one after the other, and include a heading. All figures and tables must be divided from the text.

Discussion:

The discussion is expected to be the trickiest segment to write. A lot of papers submitted to the journal are discarded based on problems with the discussion. There is no rule for how long an argument should be.

Position your understanding of the outcome visibly to lead the reviewer through your conclusions, and then finish the paper with a summing up of the implications of the study. The purpose here is to offer an understanding of your results and support all of your conclusions, using facts from your research and generally accepted information, if suitable. The implication of results should be fully described.

Infer your data in the conversation in suitable depth. This means that when you clarify an observable fact, you must explain mechanisms that may account for the observation. If your results vary from your prospect, make clear why that may have happened. If your results agree, then explain the theory that the proof supported. It is never suitable to just state that the data approved the prospect, and let it drop at that. Make a decision as to whether each premise is supported or discarded or if you cannot make a conclusion with assurance. Do not just dismiss a study or part of a study as "uncertain."



Research papers are not acknowledged if the work is imperfect. Draw what conclusions you can based upon the results that you have, and take care of the study as a finished work.

- You may propose future guidelines, such as how an experiment might be personalized to accomplish a new idea.
- Give details of all of your remarks as much as possible, focusing on mechanisms.
- Make a decision as to whether the tentative design sufficiently addressed the theory and whether or not it was correctly restricted. Try to present substitute explanations if they are sensible alternatives.
- One piece of research will not counter an overall question, so maintain the large picture in mind. Where do you go next? The best studies unlock new avenues of study. What questions remain?
- Recommendations for detailed papers will offer supplementary suggestions.

Approach:

When you refer to information, differentiate data generated by your own studies from other available information. Present work done by specific persons (including you) in past tense.

Describe generally acknowledged facts and main beliefs in present tense.

THE ADMINISTRATION RULES

Administration Rules to Be Strictly Followed before Submitting Your Research Paper to Global Journals Inc.

Please read the following rules and regulations carefully before submitting your research paper to Global Journals Inc. to avoid rejection.

Segment draft and final research paper: You have to strictly follow the template of a research paper, failing which your paper may get rejected. You are expected to write each part of the paper wholly on your own. The peer reviewers need to identify your own perspective of the concepts in your own terms. Please do not extract straight from any other source, and do not rephrase someone else's analysis. Do not allow anyone else to proofread your manuscript.

Written material: You may discuss this with your guides and key sources. Do not copy anyone else's paper, even if this is only imitation, otherwise it will be rejected on the grounds of plagiarism, which is illegal. Various methods to avoid plagiarism are strictly applied by us to every paper, and, if found guilty, you may be blacklisted, which could affect your career adversely. To guard yourself and others from possible illegal use, please do not permit anyone to use or even read your paper and file.



CRITERION FOR GRADING A RESEARCH PAPER (COMPILATION)
BY GLOBAL JOURNALS INC. (US)

Please note that following table is only a Grading of "Paper Compilation" and not on "Performed/Stated Research" whose grading solely depends on Individual Assigned Peer Reviewer and Editorial Board Member. These can be available only on request and after decision of Paper. This report will be the property of Global Journals Inc. (US).

Topics	Grades		
	A-B	C-D	E-F
<i>Abstract</i>	Clear and concise with appropriate content, Correct format. 200 words or below	Unclear summary and no specific data, Incorrect form Above 200 words	No specific data with ambiguous information Above 250 words
<i>Introduction</i>	Containing all background details with clear goal and appropriate details, flow specification, no grammar and spelling mistake, well organized sentence and paragraph, reference cited	Unclear and confusing data, appropriate format, grammar and spelling errors with unorganized matter	Out of place depth and content, hazy format
<i>Methods and Procedures</i>	Clear and to the point with well arranged paragraph, precision and accuracy of facts and figures, well organized subheads	Difficult to comprehend with embarrassed text, too much explanation but completed	Incorrect and unorganized structure with hazy meaning
<i>Result</i>	Well organized, Clear and specific, Correct units with precision, correct data, well structuring of paragraph, no grammar and spelling mistake	Complete and embarrassed text, difficult to comprehend	Irregular format with wrong facts and figures
<i>Discussion</i>	Well organized, meaningful specification, sound conclusion, logical and concise explanation, highly structured paragraph reference cited	Wordy, unclear conclusion, spurious	Conclusion is not cited, unorganized, difficult to comprehend
<i>References</i>	Complete and correct format, well organized	Beside the point, Incomplete	Wrong format and structuring



INDEX

A

Agile · 3, 2, 1
Asynchronous · 14

B

Bidirectional · 27
Breaching · 13

C

Capillaries · 15, 16, 17, 18
Chunks · 14
Concurrent · 14

E

Embedded · 26, 2
Encapsulated · 15, 17, 18, 27
Encoder · 17
Erratic · 30, 35, 36

I

Implant · 3
Ingests · 2
Integration · 20, 3, 2, 3, 2, 1, 2

M

Multicore · 21

N

Neurons · 18

P

Pandemic · 1
Practitioners · 27, 34, 35

R

Rumors · 1, 2, 11

S

Satirical · 2
Sigmoid · 3

T

Trauma · 1

U

Unidirectional · 2



save our planet



Global Journal of Computer Science and Technology

Visit us on the Web at www.GlobalJournals.org | www.ComputerResearch.org
or email us at helpdesk@globaljournals.org



ISSN 9754350