APEX ACADEMIA PRESS
ELEGANCE ON EVERY RESEARCH

International Journal of Machine Learning, AI & Data Science Evolution
ISSN: 3068-5073/ Open Access / IJMLAIDSE_Y1F1A002
Apex Academia Press / https://doi.org/10.63665/ijmlaidse-y1f1a002

Original Article

Deep Learning-Based Email Spam Identification and Classification for Enhanced Cybersecurity

Vaibhav Maniar[1], Aniruddha Arjun Singh Singh[2], Rami Reddy Kothamaram[3], Dinesh Rajendran[4], Venkata Deepak Namburi[5], Vetrivelan Tamilmani[6]

[1]*Oklahoma City University, MBA / Product Management, USA.*
[2]*ADP, Agile Team Leader. USA.*
[3]*California University of management and science, MS in Computer Information systems, USA.*
[4]*Coimbatore Institute of Technology, MSC. Software Engineering, USA.*
[5]*Department of Computer Science, University of Central Missouri, USA.*
[6]*Principal Service Architect, SAP America*

Abstract

Digital communication and cybersecurity are major problems concerning the spread of unsolicited and malicious emails. Spam emails are vectors of phishing, malware and financial fraud as well as inbox clutter. The conventional spam detection methods, like rule-based spam filters and the classical machine learning (ML) models, have limitations on their ability to detect spam based on pre-determined patterns, feature engineering, which requires a lot of human effort and inability to adapt to changes in threats. To address these, the given piece uses a deep learning-driven framework that would allow the extraction of complex and hierarchical patterns in email data automatically. Model training and evaluation are performed using the Spam Base benchmark dataset comprising 4601 emails having 57 features. High-quality input data can be guaranteed by doing comprehensive preprocessing such as parsing, tokenization, stemming, case folding, error correction, and extraction with the help of regressions. Feature extraction, dimensionality reduction, and data classification selection techniques are used to further optimize the dataset. Accuracy, Recall, Precision, and F1-score were evaluated using a 70:30 train-test split for an Artificial Neural Network (ANN), with results of 99.50, 99.68, 99.68, and 99.68, respectively. The findings prove the strength, scalability, and usefulness of the framework in spam detection, which helps to increase cybersecurity and efficient email communication systems. Further improvements on the detection of advanced spamming can be discussed in future work, and one of the methods is the multi-mode and transformer-based approach.

Keywords

Email Spam Detection, Deep Learning, Machine Learning Cybersecurity, Malicious Email Filtering, Email Classification, Support Vector Machines (SVM), Naive Bayes (NB). Long Short-Term Memory (LSTM).

Article History

Received:
24.01.2025

Accepted:
25.02.2025

Published:
15.03.2025

1. Introduction

The emergence of email as a powerful form of online communication has changed how individuals, businesses, and organisations communicate [1]. Email, as a cheap and easy method of communication, was easily adopted as a necessary means of communication both at the professional and personal levels. It has however been exploited by the fact that it has become very common [2]. The abuse of email has over time come up with unsolicited and irrelevant messages commonly referred to as spam. These spam emails are of bulk advertisements, fake marketing campaigns to a malicious intent of phishing, malware distribution and financial fraud [3]. Although spam is considered a simple inconvenience to some users, its extended effects are loss of productivity, information congestion, breach of data and even a significant economic effect to a large scale [4].

Spam has also been in a constant state of development, implementing some countermeasures to detection techniques and becoming more advanced. In the previous stages, spam was easily spotted and blocked using basic filters [5]. Attackers today develop fraudulent messages that are very similar to authentic messages hence making them very difficult to detect. Email has therefore turned into a pivotal point of vulnerability to cybersecurity attacks where it acts as a medium of phishing attacks, ransomware attacks and identity theft [6]. Due to the ever-growing use of digital communication, the need to have solid email security has not only become convenient but a strong necessity in the protection of sensitive data and the safeguard of organisations against cybercrime [7].

In an attempt to overcome these obstacles, conventional techniques of rule-based filters, blacklists, whitelists and Bayesian filters were first introduced [8]. Although they are effective to some extent, such methods are based on static patterns and a set of present rules, and attackers can readily circumvent them [9]. These shortcomings led to the use of ML, which enabled it to perform automated recognition through learning patterns in email data. The algorithms that were reasonable enough to identify spam were SVM, NB, and random forests. Nevertheless, ML-based models are not without major challenges, they need large amounts of manual feature engineering, do not support unstructured or big data, and are not flexible to novel and changing spam strategies [10]. Deep learning (DL) has proved to be a more potent fix to such issues. When compared to the traditional ML algorithms, deep learning models are capable of automatically extracting complex and hierarchical features of raw email data without any manual process [11][12]. Convolutional Neural Networks (CNNs) are useful in finding patterns that are not easily visible, whereas Recurrent (RNNs) and LSTM models are more efficient with sequential data, which is especially effective with text analysis. More recently, the transformer-based models have improved spam detection by including more profound semantic relationships in emails [13].

Not only these models enhance accuracy but they are also more adaptive to changing spam techniques [14].Using the power of DL, classification and email spam detection can transcend the limitations of older methods, providing scalable, flexible and very accurate solutions. This change is critical to improving the cybersecurity of the Internet, to protecting Web users, and to the further viability of email as a reliable form of online communication.

A. Motivation and Contribution of the Paper

The motivation for this work arises from the rapid growth of email communication and the parallel increase in spam emails that pose significant risks to cybersecurity. As well as cluttering inboxes, spam can be used in the distribution of malware and financial fraud, as well as phishing. Conventional spam detection methods based on rule-based spam filters or shallow machine learning models tend to be less accurate and do not scale as well in order to counter emerging spam strategies. To overcome them, this paper will apply deep learning techniques which have potential to automatically derive intricate trends in textual information to attain a robust and adaptive spam classifications. The suggested solution will increase the accuracy of detection, which will reduce cases of false classification, and thus help to promote secure and effective communication via email. The most important findings of this work are as follows:

- There, the benchmark data set Spam Base was used with 4,601 email samples (spam) having 57 spam-related engineered features.
- Conducted extensive preprocessing, such as, parsing, tokenization, stemming, case folding, error correction and extracting features using regular expressions.
- Semi-automated extraction of features, dimensionality reduction and feature selection to narrow down the input data to train a model.
- Trained a model of ANN and implemented it to be effective in classifying emails into ham and spam.
- Assessed the model on basis of F1-score, recall, precision, and accuracy, maintaining equal performance levels on all of the metrics.
- Illustrated the relevance of the DL-based models in improving email spam detection and offering better cybersecurity defence solutions.

B. Justification and Novelty of the Paper

The increased rates of spam emails are major threats to digital communication and cybersecurity that require sophisticated methods of detection. The problem with traditional and conventional methods of machine learning is they are rule-based and thus fail to evolve in line with the spam strategies, making them less effective and less scalable. The originality of the research is utilize DL to identify complex and hierarchical patterns in email data automatically, without using manual feature engineering. By integrating comprehensive preprocessing, feature optimization, and a robust neural network framework, the approach enhances adaptability and accuracy in email classification. This study provides a scalable, flexible, and effective solution for contemporary spam detection challenges.

C. Structure of Paper

The following is the paper's structure: The second section discusses relevant research on detecting spam in email. Section III outlines the methodology, taking into account the features engineered and pre-processing procedures performed on the dataset. Section IV presents the experimental setup, outcomes, and assessment of performance along with a review of results. Lastly, Section V summarizes key outputs and concludes the paper, emphasizing main contributions, and suggesting future research directions for enhanced cybersecurity applications.

2. Literature Review

Most of the current literature on email spam detection uses deep learning and machine learning to get very accurate results; however, gaps remain in multi-modal detection, large-scale evaluations, and hybrid deep learning approaches for enhanced cybersecurity.

Alauthman (2020) highlights the significant issue of email spam, which is a growing problem originating from botnets worldwide. Time spent identifying spam emails, the safety of personal mail, and mailbox capacity are all impacted. A Gated Recurrent Unit Recurrent Neural Network (GRU-RNN) with SVM was created to recognise bot spam emails in order to solve this problem. The method achieved a 98.7 per cent success rate when tested on the Spambase dataset, demonstrating its excellent capability in detecting spam emails. [15]

Rahman and Ullah (2020) propose a novel approach to spam message detection that uses sentiment analysis of email body content. They use a bidirectional LSTM network and word embeddings to assess texts' sequential and emotional features. Higher-level text features are extracted by the model using a Convolution Neural Network. With an improved accuracy of about 98-99%, the model outperforms both popular machine learning classifiers and state-of-the-art approaches for spam message detection. This model is evaluated using f-score, recall, and precision on two datasets: Ling spam and spam text message categorisation [16].

Khamis et al. (2019) conducted a study aiming to determine possible email header characteristics to use spam in two email datasets: Cybersecurity Data Mining on Sites and Anomaly Detection Difficulties. The main objective was to use SVM with Weka 3.9.2 and RapidMiner Studio to produce relevant features and classify them. In this approach, there were five steps gathering data, selecting features, pre-processing data, classification, and detection. With accuracy rates of 88.80% and 87.20%, respectively, the SVM classifier outperformed the Anomaly Detection Challenges dataset.[17]

Akinrele's (2019) study indicates how the issue of email spam and phishing has been on the rise and spammers and spammers are ever-inventing methods to overcome the current software. Filtering is a popular method for removing spam and phishing, while phishing detection relies on validating email body and URLs. In order to decrease feature space dimensionality and increase accuracy, the paper suggests an ensemble approach for feature selection techniques based on spam and phishing filters. The study used Machine Learning-based mRMR models and Ensemble models, resulting in an average 83% accuracy for seven classifiers. This feature selector could potentially legitimize future email cyber-attacks, indicating the potential for further research and expansion.[18]

Göker (2018) highlights the importance of detecting spam e-mails, which are fake, falsified emails aiming to collect sensitive personal information or act against authority illegally. The majority of emails have spam or relevant spam-like content, such phishing emails. It is essential to identify these emails in order to stop illegal access to user credentials. Effective machine learning and classification techniques are necessary for prompt processing in order to identify spam emails. With billions of emails on the internet, automatic classification of emails as spam or not is an important problem. Research on supervised machine learning, more especially "deep learning" techniques, has shown encouraging results in successfully classifying emails with up to 96% accuracy. [19]

Kumar et al. (2018) used the Hidden Markov Model and the ID3 algorithm to detect spam emails. Spam emails are a serious problem since they waste transmission bandwidth, memory, money, and time. By computing the overall likelihood of an email using all posteriorly categorised words in emails, the model classifies emails as either ham or spam. For this investigation, the Enron dataset of 5172 emails of which 2086 were pre-classified as spam and 2086 as ham was used. The significance of handling spam emails in the communication process was underscored by the testing results, which demonstrated an accuracy of 89% on spam emails. [20]

Table 1 summarizes key studies on email spam detection, detailing methodologies, datasets, findings, limitations, and future directions, and highlights research gaps in deep learning approaches, multi-modal detection, and scalable cybersecurity solutions

Table 1: Summary of Existing Research on Email Spam Detection and Classification

Author(s)	Methodology	Dataset	Key Findings	Limitations	Future Direction
ALAUTHMAN (2020)	GRU-RNN with SVM	Spam-base	Achieved 98.7% accuracy; effective in detecting spam emails	Only tested on a single dataset; limited comparison with other deep learning models	Explore multi-dataset evaluation; hybrid deep learning architectures
Rahman & Ullah (2020)	Sentiment Analysis + Word Embeddings + Bi-LSTM + CNN	LingSpam, Spam Text Message Classification	Accuracy 98-99%; outperformed traditional ML classifiers	Focuses on textual sentiment; may not generalize to non-textual/spam formats	Extend to multi-modal spam detection (attachments, images, links); optimize training efficiency
Khamis et al. (2019)	SVM on email header features	Anomaly Detection Challenges, CSDM2010	SVM achieved ~88% accuracy; header features useful for classification	Limited to header features; no deep learning approach	Integrate header + content features with deep learning models
Akinrele (2019)	Ensemble ML + mRMR feature selection	Public email datasets	Average accuracy 83%; reduced feature dimensionality; improved spam & phishing classification	Accuracy lower than deep learning models; limited dataset variety	Explore deep learning ensembles and larger datasets for higher accuracy
Göker, (2018)	Supervised Deep Learning	Email datasets	Deep learning effective; up to 96% accuracy in email classification	may lack generalizability	Evaluate on large-scale datasets; apply advanced DL models like transformers
Kumar et al., (2018)	ID3 Decision Tree + HMM	Enron Dataset (5172 emails)	Achieved 89% accuracy; combined probabilistic & decision tree approaches	Outdated dataset; limited spam variety; non-DL approach	Test deep learning approaches; use larger, more diverse datasets

3. Methodology

This methodology presents a DL-based framework for email spam detection and classification to enhance cybersecurity by reducing malicious and irrelevant messages as shown in figure 1. The Spam Base benchmark dataset was utilized, containing 4,601 email records, with 1,813 (39%) labelled as spam and 2,788 (61%) as ham. Each record is described by 57 features, including word frequencies, character frequencies, and statistics on capital letter sequences, making it suitable for spam classification tasks. The methodology begins with a comprehensive data preparation phase involving parsing, tokenization, stemming, case folding, error correction, and regex-based extraction to ensure consistency and quality. Feature extraction is then performed to capture discriminative patterns, followed by dimensionality reduction and feature selection to optimize model learning. Testing (30%) and Training (70%) sets make up the dataset. Using the processed data, a suggested ANN model is trained, and the F1-score, accuracy, recall, and precision are used to evaluate its effectiveness. This systematic approach ensures reliable and scalable spam filtering for real-world applications.

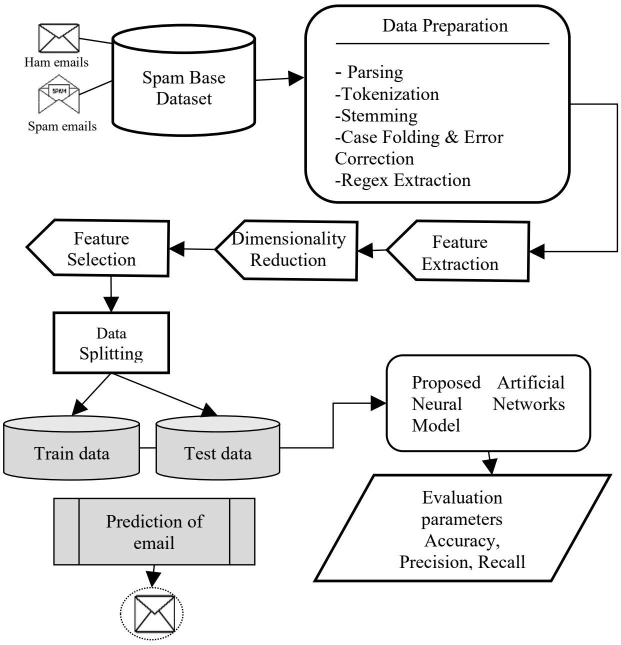

Fig-1: Methodology Framework for Deep Learning-Based Email Spam Classification and Detection

A. Data Collection

There are 4,601 email messages in the Spam Base collection, with 1,813 rows labelled as spam (39% of the total) and 2,788 rows as ham (61%). For every email in the sample, a total of 57 characteristics are retrieved. Some distinguishing terms' frequencies in the message body are depicted by these characteristics. Among its features are 48 real-numerical ones that denote the frequency of words like "remove," "address," "order," "internet," "receive," "mail," "business," "free," "credit," "money," "data," and longest, "meeting." The average and the overall length of subsequent capital-letter sequences are quantified by features 55–57. Finally, there is class 1 for spam and class 0 for valid emails. This is what the data visualisations look like:

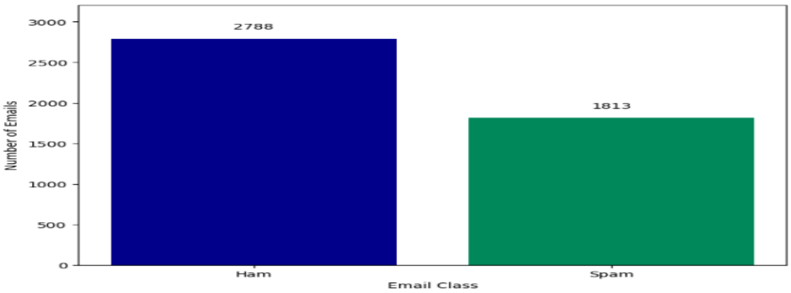

Fig-2: Distribution of Spam vs. Ham Emails

Figure 2 is a bar chart that displays the count of two different email classifications: Ham (legitimate emails) and Spam (unwanted or junk emails). The chart shows two vertical bars. The bar for Ham emails is coloured dark blue and is significantly taller than the Spam bar. A number is positioned at the top of each bar, indicating the exact count. There are 2788 emails classified as Ham and 1813 emails classified as Spam. This visualization effectively compares the number of emails in each category, illustrating that in this dataset, there are more legitimate emails than spam emails.

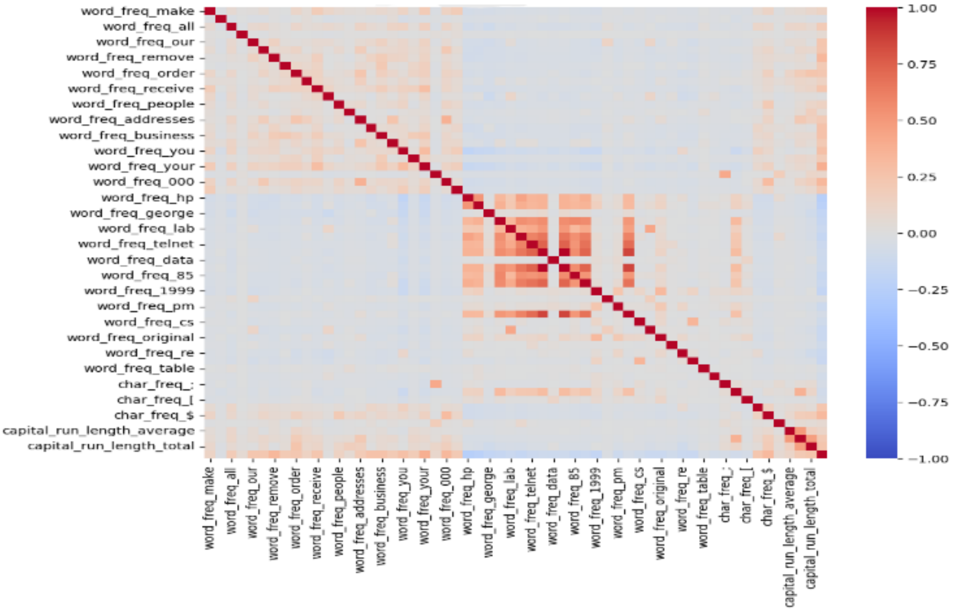

Fig-3: Correlation Heatmap of Features in the Spambase Dataset

A correlation heatmap for the Spambase dataset's features is shown in Figure 3, illustrating the pairwise relationships between 57 attributes related to email characteristics. The diagonal elements, shown in deep red, indicate perfect positive correlation (correlation = 1) of each feature with itself. Shades of red represent positive correlations between different features, while shades of blue indicate negative correlations. Most features show weak correlations, while certain groups, such as specific word frequency features (e.g., word_freq_george, word_freq_edu) and character frequency features (e.g., char_freq_!), display moderate correlations. This visualisation helps identify redundant or highly interdependent features for modelling.

B. Data Preparation

The data preparation stage focuses on transforming raw email collections into a uniform and analysable structure. This process ensures consistency, eliminates irregularities, and organizes the dataset with appropriate labels for reliable training and evaluation. The following pre-processing steps were performed:

- Emails are parsed to separate header, subject, body, and sender information in a structured format.
- Splitting text into smaller pieces, such as words, is one method for easier analysis.
- Words are reduced to their root or base form through stemming and lemmatization.
- All characters are converted to lowercase to maintain consistency.
- Spelling and typographical mistakes are corrected using similarity scoring.
- Regular expressions are applied to detect and extract URLs, domain names, and suspicious keywords.

C. Normalization

The numerical characteristics that have been created out of the emails, like word counts, character counts, and link counts, are put to scale to make sure that they contribute equally to the learning process [21]. To put values on a common range, scaling and quantization are applied to put the values in a standard range, usually [0,1], which removes biasness of attributes with bigger magnitudes and stabilizes the training. Min-Max normalization is the most widespread method, as it is expressed by eq(1):

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (1)$$

x is the starting value, x^' is the scaled value, and x_max and x_min are the feature's maximum and minimum values.

D. Feature Extraction

The process of transforming raw and cleaned email data into measurable attributes is known as Feature extraction which can be fed into the classification model. In this study, forty features are extracted and grouped into five categories that describe different aspects of an email. Body-based features include HTML tags, forms, suspicious keywords, and text statistics. Subject-based features account for message length, reply/forward indicators, and terms such as "verify" or "bank." Sender address features examine address length and domain validity. URL-based features capture link counts, port usage, and special characters, while script-based features detect JavaScript, pop-ups, and on-click events.

E. Dimensionality Reduction

Dimensionality reduction aims to retain the most pertinent data while decreasing the feature space's variable count [22]. The study uses PCA to separate associated features into independent principal components. By retaining solely the most highly variable components, PCA ensures that the essential patterns in the data are preserved, while irrelevant variations are discarded. This reduces the dimensionality of the dataset, lowers computational cost, and improves training efficiency. At the same time, it helps the neural network focus on the most informative characteristics of the emails, boosting generalization and overall performance.

F. Feature Selection

The term "feature selection" refers to the steps used to determine which features are most useful for categorisation and then retain only those, while removing those with little or no impact. This step helps reduce model complexity and enhances performance. In this work, two techniques are applied. Low Variance Filtering discards features that show minimal variation across email classes, as they carry little discriminative value. The Chi-Squared Test, on the other hand, statistically measures the dependence between each feature and the target class, retaining only the most relevant ones. These approaches ensure that the classifier focuses on strong predictors for accurate detection.

G. Data Splitting

The dataset is split up into training and testing subsets to ensure proper model evaluation. There is a 70:30 split, meaning that 70% of data is utilised for training the classifier and 30% is reserved for validating and testing its performance.

H. Proposed Artificial Neural Networks Model

The original intent of creating Artificial Neural Networks (ANN) was to create a computer model that could mimic brain activity. Drawing on the inner workings of specific cells in the brain known as neurones, it is a computer simulation of a neural network. [23]. There is some theoretical support for the idea that ANN can learn concepts just like the human brain. An artificial neural network (ANN) fundamental building block is the neuron or node. Through the weighted edges or connections, the input is received by other nodes. [24]. The inputs' relative importance determines the connection weights. The output is linked to the sum of the inputs that are weighted. The activation function, f(x), and equation (2) that calculates the output, y_1, are:

$$y_1 = f(w_1x_1 + w_2x_2 + \cdots + w_nx_n + b) \qquad (2)$$

Here, $x_1, x_2, \dots x_n$ Regarding the inputs, $w_1, w_2, \dots w_n$ Regarding the weights, y_1 is the output of the neuron, "b" denotes the bias, while "f(ㅈ)" denotes the activation function. At output layer, the network combines the results of previous hidden layers to produce final prediction. This can be represented by Equation. (3)

$$y = f^{(L)}\left(W^{(L)}a^{(L-1)} + b^{(L)}\right) \qquad (3)$$

The vector of outputs from the last hidden layer is represented by a^((L-1)), the weights and biases of the output layer are denoted as W^((L)) and b^((L)), the activation function is denoted as f^((L)) and the final output of the network is equal to y.

I. Evaluation Parameters

F1 score, recall, precision, and accuracy are some of the criteria used to evaluate the outcomes. Table 2s confusion matrix serves as the basis for the computation of these matrices. A two-by-two matrix is appropriate for a binary classification issue. The real class labels are on top, while the projected ones are along the side. In the matrix, you can see how many predictions the classifier made for each cell's category [25]. Various labels of matrix are defined as:

- TN: Negative labels predicted as negative
- TP: Positive labels predicted as positive
- FN: Positive labels wrongly predicted as negative
- FP: Negative labels wrongly identified as positive

Table 2: Confusion matrix

Type	Ham	Spam
Ham	True Positive (TP)	False Positive (FP)
Spam	False Negative (FN)	True Negative (TN)

The following is the formula for calculating the various evaluation metrics:

Accuracy: To calculate accuracy, divide by the total number of correct predictions in both classes, and find the total number of guesses. To get the percentage, it is multiplied by 100. The computation is illustrated in Equation (4)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (4)$$

Precision: Equation (5) provides a mathematical representation of precision p, It is described as the method of determining the precise and, on occasion, the predictive value of a more favourable result [26].

$$Precision = \frac{TP}{TP+FP} \qquad (5)$$

Recall: The ratio of the total number of objects recalled to the measure of completeness is called recall, which is also called responsiveness [27]. When divided by the entire number of records in the database, it gives an average of the number of records retrieved. Equation (6) is used to compute it:

$$Recall = \frac{TP}{TP+FN} \qquad (6)$$

F1-Score: Since F1 scores are a weighted average of recall and precision, both are taken into account while determining them. Its value ranges from 0 (the worst case scenario) to 1 (the best case scenario). Equation (7) shows the calculation:

$$F1\ score = \frac{2.(Precision \cdot Recall)}{Precision+Recall} \qquad (7)$$

These evaluation measures offer a thorough analysis of model performance, guaranteeing forecasts that are strong, precise, and dependable.

4. Results Analysis and Discussions

The suggested model for email spam detection and categorisation was built using an ANN. The implementation was done in Python to be efficient and scalable and the statistical data analysis was done using the WEKA tool. The experiments were implemented on a personal computer based on Ubuntu 20.04.1 LTS and using AMD A6 processor (2.6 GHz) and 16 GB RAM. To train the neural network, the calculation was moved to Google Colab GPU using TensorFlow as the backend of the Keras model to streamline developing and running models. In order to evaluate the model's efficacy, typical measures of classification, such as F1-score, Recall, Precision, Accuracy, and which are essential in determining the strength of spam detection systems, have been used. Table 3 presents the results.

Table 3: Evaluation Results of ANN for Email Spam Detection

Parameter	ANN
Accuracy	99.50
Precision	99.68
Recall	99.68
F1-score	99.68

The model had the ability to detect spam and legitimate emails with an accuracy 99.50% of 99.50% with a minimal error rate. The ANN has low false positives and negatives with a single recall, precision, and F1-score of 99.68, but it is also balanced. The almost flawless result indicates the effectiveness of the deep learning, strong feature representation, and training on the basis of the use of the GPUs, which demonstrates the trustworthiness of the ANN in terms of improving the level of cybersecurity by detecting spam.

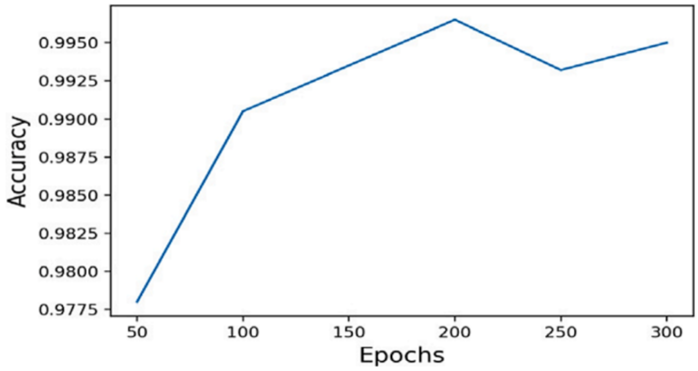

Fig-4: Accuracy Curve of the ANN Model Across Epochs

Figure 4 represents a line graph by which the performance of an ANN model can be examined with regard to the identification of email spam. The graph will be against the x-axis of the number of Epochs and the y-axis of Accuracy. The accuracy which is a percentage of correct prediction generally increases with the number of epochs the model is trained. It catches up on the levels of about 0.9775 at 50 epochs and peaks at an accuracy of about 0.996 at 200 epochs. Nonetheless, following this peak, the precision is negligible, and then it rises. This is an indication that the model is a good learner but it could have slight variations in the performance during the course of training.

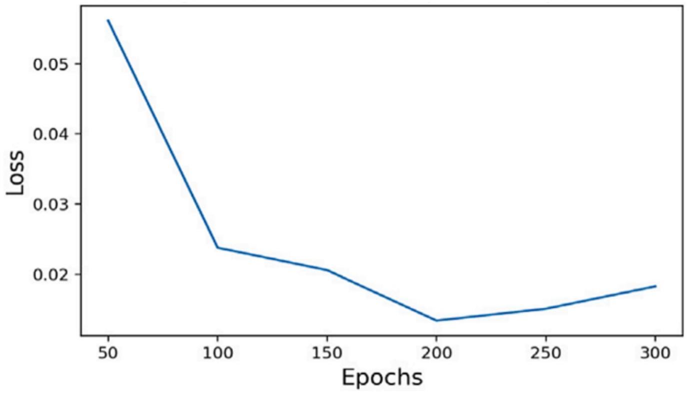

Fig-5: Loss Curve of the ANN Model Across Epochs

Figure 5 is a line graph that presents the association between "Epochs" and "Loss" of a model of ANN which is employed to detect email spam. The number of training epochs, 50-300, is plotted on the x-axis and the value of the loss, an expression of the error committed by the model, is plotted on the y-axis. The graph shows that the loss is likely to reduce with the expansion of the epoch count. Nonetheless, at a certain point in the curve, at about 200 epochs, the model may begin to overfit the training data if the loss begins to rise marginally. This plot is an essential device in the control over the process of training and in the optimization of the performance of the model.

A confusion matrix to performance of an ANN model to detect email spam is shown in Figure 6. The two axes of the matrix are "True Class" and "Predicted Class" where both have the categories of Ham and Spam. In the upper-left cell, there are 3715 emails that were correctly marked as True Negatives (Ham). The upper-right cell means that 20 emails were Ham but had been mistaken in the category as Spam (False Positives). The cell at the bottom-left represents emails that were 35 in total Spam yet they were classified as Ham (False Negatives). Lastly, the cell in the bottom-right corner depicts 6230 emails that were rightly labelled as being Spam (True Positives). This matrix is very important in summarizing accuracy of model and as it has been observed a lot of correct predictions and few misclassifications.

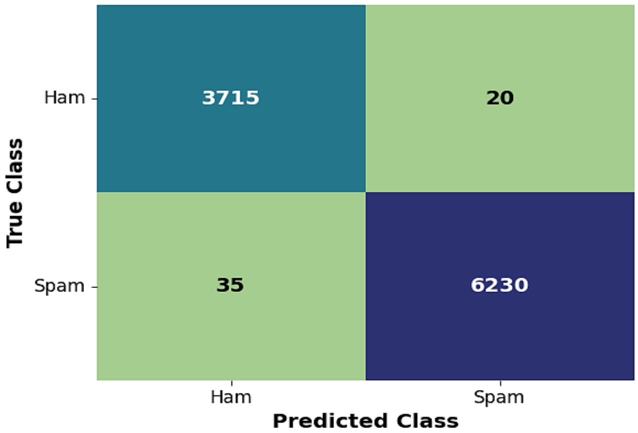

Fig-6: Confusion Matrix of ANN-Based Spam Detection Model

A. Comparative Analysis

The work of the proposed ANN-based deep learning model in detecting email spam was against other models, such as a standard Neural Network, MLP (Multilayer Perceptron), and LSTM. Table 4 is a summary of the findings.

Table 4: Comparative Assessment of Email Spam Detection Models Using Spambase Dataset

Models	Accuracy	Precision	Recall	F1-score
Neural network[28]z	91.8	90.5	88.1	89.3
MLP[29]	92.3	92.3	92.3	92.3
LSTM[30]	92.93	93.58	98.22	95.84
ANN	99.50	99.68	99.68	99.68

It is clear that ANN model is superior to others. The default Neural Network demonstrates the worst performance, its accuracy is 91.8, and its F1-score is 89.3, which means that it is not quite efficient in spam detection. MLP is better in this regard with regular measures of 92.3 with moderate but steady improvements. LSTM has a better recall of 98.22, which is effective in the detection of most spam but even the precision (93.58) and F1-score (95.84) are worse than ANN, implying that LSTM generates a larger number of false positives than ANN does. The high metrics of the ANN in all measures show its ability to precisely classify emails, reduce false positive and false negative emails, and learn complicated trends in email statistics successfully. This renders it a very strong and sound option to improve cybersecurity concerning spamming.

5. Conclusion and Future Work

The deep learning-based model temperature in this paper illustrates a very efficient methodology to classify and detect email spam and eliminate the weaknesses of the conventional rule-based and ML methods. With the help of an ANN trained based on the Spam Base dataset, the model demonstrated very good results, with a 99.68% F1-score, 99.68% Recall, 99.68% Precision, and 99.50% Accuracy, demonstrating good performance in the field of spam and legitimate emails. The preprocessing, feature extraction, dimensionality reduction and feature selection steps made sure that the input data was of good quality so that the ANN could learn the complex patterns and dependencies that the traditional methods could have overlooked. The ANN was considered to have a better capability to reduce the false positives and negatives and at the same time balance performance across the evaluation measures when compared to the standard neural networks, MLP and LSTM models. The next step in development of model is consideration of multi-modal spam detection model which adds to the text-only model attachments, pictures, and embedded links to counter more complex threats. The addition of transformer-based architecture may also promote semantic interpretation and flexibility to the changing spam tactics. Also, the implementation of the system in the large-scale enterprise email infrastructures will test scalability, real-time performance, and robustness. Regular updates of datasets and the dynamic process of learning will assist in keeping the models relevant, which would guarantee further improvement of cybersecurity and stable email communication.

6. References

[1] V. Rajavel, G. Balaji, and A. V. Gomathinayagam, "Eye Gaze Pecularities Detection in Children with Autism using a Head-free cam," Int. J. Eng. Sci. Res. Technol., vol. 5, no. 6, pp. 868–876, 2016.

[2] A. Zamir, H. U. Khan, W. Mehmood, T. Iqbal, and A. U. Akram, "A feature-centric spam email detection model using diverse supervised machine learning algorithms," Electron. Libr., vol. 38, no. 3, pp. 633–657, 2020, doi: 10.1108/EL-07-2019-0181.

[3] Y. Alamlahi and A. Muthana, "An Email Modelling Approach for Neural Network Spam Filtering to Improve Score-based Anti-spam Systems," Int. J. Comput. Netw. Inf. Secur., vol. 10, no. 12, pp. 1–10, Dec. 2018, doi: 10.5815/ijcnis.2018.12.01.

[4] D. D. Rao, "Multimedia Based Intelligent Content Networking for Future Internet," in 2009 Third UKSim European Symposium on Computer Modeling and Simulation, 2009, pp. 55–59. doi: 10.1109/EMS.2009.108.

[5] E. E. Eryılmaz, D. Ö. Şahin, and E. Kılıç, "Filtering Turkish Spam Using LSTM From Deep Learning Techniques," in 2020 8th International Symposium on Digital Forensics and Security (ISDFS), IEEE, Jun. 2020, pp. 1–6. doi: 10.1109/ISDFS49300.2020.9116440.

[6] S. S. S. Neeli, "Real-Time Data Management with In-Memory Databases: A Performance-Centric Approach," J. Adv. Dev. Res., vol. 11, no. 2, p. 8, 2020.

[7] H. Bhuiyan, A. Ashiquzzaman, T. I. Juthi, S. Biswas, and J. Ara, "A survey of existing e-mail spam filtering methods considering machine learning techniques," Glob. J. Comput. Sci. Technol., vol. 18, no. 2, pp. 20–29, 2018.

[8] S. Bosaeed, I. Katib, and R. Mehmood, "A Fog-Augmented Machine Learning based SMS Spam Detection and Classification System," in 2020 Fifth International Conference on Fog and Mobile Edge Computing (FMEC), IEEE, Apr. 2020, pp. 325–330. doi: 10.1109/FMEC49853.2020.9144833.

[9] Y. R. Bujang and H. Hussin, "Should we be concerned with spam emails? A look at its impacts and implications," in 2013 5th International Conference on Information and Communication Technology for the Muslim World (ICT4M), 2013, pp. 1–6.

[10] A. Balasubramanian, "Proactive Machine Learning Approach to Combat Money Laundering in Financial Sectors," Int. J. Innov. Res. Eng. Multidiscip. Phys. Sci., vol. 7, no. 2, pp. 1–15, 2019, doi: 10.5281/zenodo.14508474.

[11] D. Wang, D. Irani, and C. Pu, "A Study on Evolution of Email Spam Over Fifteen Years," in Proceedings of the 9th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing, ICST, 2013, pp. 1–10. doi: 10.4108/icst.collaboratecom.2013.254082.

[12] A. Tyagi, "Content based spam classification-a deep learning approach," Univ. Calgary Calgary, AB, Canada, 2016.

[13] A. Balasubramanian and N. Gurushankar, "AI-Driven Supply Chain Risk Management : Integrating Hardware and Software for Real-Time Prediction in Critical Industries," Int. J. Innov. Res. Eng. Multidiscip. Phys. Sci., vol. 8, no. 3, 2020, doi: 10.5281/zenodo.14565873.

[14] S. S. S. Neeli, "Optimizing Database Management with DevOps: Strategies and Real-World Examples," J. Adv. Dev. Res., vol. 11, no. 1, 2020.

[15] M. Alauthman, "Botnet Spam E-Mail Detection Using Deep Recurrent Neural Network," Int. J. Emerg. Trends Eng. Res., vol. 8, no. 5, pp. 1979–1986, 2020, doi: 10.30534/ijeter/2020/83852020.

[16] S. E. Rahman and S. Ullah, "Email Spam Detection using Bidirectional Long Short Term Memory with Convolutional Neural Network," in 2020 IEEE Region 10 Symposium (TENSYMP), 2020, pp. 1307–1311. doi: 10.1109/TENSYMP50017.2020.9230769.

[17] S. A. Khamis, C. F. M. Foozy, M. F. A. Aziz, and N. Rahim, "Header Based Email Spam Detection Framework Using Support Vector Machine (SVM) Technique," in International conference on soft computing and data mining, 2020, pp. 57–65. doi: 10.1007/978-3-030-36056-6_6.

[18] M. O. Akinrele, "Detection of Phishing and Spam Emails Using Ensemble Technique," Dublin, National College of Ireland, 2019.

[19] O. Göker, "Spam filtering using big data and deep learning," 2018.

[20] V. Kumar, Monika, P. Kumar, and A. Sharma, "Spam Email Detection using ID3 Algorithm and Hidden Markov Model," in 2018 Conference on Information and Communication Technology (CICT), 2018, pp. 1–6. doi: 10.1109/INFOCOMTECH.2018.8722378.

[21] Y. K. Zamil, S. A. Ali, and M. A. Naser, "Spam image email filtering using K-NN and SVM," Int. J. Electr. Comput. Eng., vol. 9, no. 1, pp. 245–254, 2019, doi: 10.11591/ijece.v9i1.pp245-254.

[22] T. Wu et al., "Detecting spamming activities in twitter based on deep-learning technique," Concurr. Comput. Pract. Exp., vol. 29, no. 19, Oct. 2017, doi: 10.1002/cpe.4209.

[23] G. Jain, M. Sharma, and B. Agarwal, "Spam detection on social media text," Int. J. Comput. Sci. Eng., vol. 5, 2017.

[24] S. Srinivasan et al., “Deep Convolutional Neural Network Based Image Spam Classification,” in 2020 6th Conference on Data Science and Machine Learning Applications (CDMA), IEEE, Mar. 2020, pp. 112–117. doi: 10.1109/CDMA47397.2020.00025.

[25] H. M, N. A. Unnithan, V. R, and S. KP, “Deep learning based phishing e-mail detection,” in Proc. 1st AntiPhishing Shared Pilot 4th ACM Int. Workshop Secur. Privacy Anal.(IWSPA), 2018, pp. 1–5.

[26] D. Mallampati and N. Hegde, “A Machine Learning Based Email Spam Classification Framework Model: Related Challenges and Issues,” Int. J. Innov. Technol. Explor. Eng., vol. 9, no. 4, pp. 3137–3144, 2020, doi: 10.35940/ijitee.d1561.029420.

[27] S. Madisetty and M. S. Desarkar, “A neural network-based ensemble approach for spam detection in Twitter,” IEEE Trans. Comput. Soc. Syst., vol. 5, no. 4, pp. 973–984, 2018.

[28] S. S. Roy, A. Sinha, R. Roy, C. Barna, and P. Samui, “Spam Email Detection Using Deep Support Vector Machine, Support Vector Machine and Artificial Neural Network,” in International Workshop Soft Computing Applications, 2018, pp. 162–174. doi: 10.1007/978-3-319-62524-9_13.

[29] S. M. Abdulhamid, M. Shuaib, and O. Osho, “Comparative analysis of classification algorithms for email spam detection,” 2018.

[30] G. Jain, M. Sharma, and B. Agarwal, “Spam detection in social media using convolutional and long short term memory neural network,” Ann. Math. Artif. Intell., vol. 85, no. 1, pp. 21–44, 2019.

[31] Polam, R. M., Kamarthapu, B., Kakani, A. B., Nandiraju, S. K. K., Chundru, S. K., & Vangala, S. R. (2021). Big Text Data Analysis for Sentiment Classification in Product Reviews Using Advanced Large Language Models. International Journal of AI, BigData, Computational and Management Studies, 2(2), 55-65.

[32] Gangineni, V. N., Tyagadurgam, M. S. V., Chalasani, R., Bhumireddy, J. R., & Penmetsa, M. (2021). Strengthening Cybersecurity Governance: The Impact of Firewalls on Risk Management. International Journal of AI, BigData, Computational and Management Studies, 2, 10-63282.

[33] Pabbineedi, S., Penmetsa, M., Bhumireddy, J. R., Chalasani, R., Tyagadurgam, M. S. V., & Gangineni, V. N. (2021). An Advanced Machine Learning Models Design for Fraud Identification in Healthcare Insurance. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 2(1), 26-34.

[34] Kamarthapu, B., Kakani, A. B., Nandiraju, S. K. K., Chundru, S. K., Vangala, S. R., & Polam, R. M. (2021). Advanced Machine Learning Models for Detecting and Classifying Financial Fraud in Big Data-Driven. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 2(3), 39-46.

[35] Tyagadurgam, M. S. V., Gangineni, V. N., Pabbineedi, S., Penmetsa, M., Bhumireddy, J. R., & Chalasani, R. (2021). Enhancing IoT (Internet of Things) Security Through Intelligent Intrusion Detection Using ML Models. International Journal of Emerging Research in Engineering and Technology, 2(1), 27-36.

[36] Vangala, S. R., Polam, R. M., Kamarthapu, B., Kakani, A. B., Nandiraju, S. K. K., & Chundru, S. K. (2021). Smart Healthcare: Machine Learning-Based Classification of Epileptic Seizure Disease Using EEG Signal Analysis. International Journal of Emerging Research in Engineering and Technology, 2(3), 61-70.

[37] Kakani, A. B., Nandiraju, S. K. K., Chundru, S. K., Vangala, S. R., Polam, R. M., & Kamarthapu, B. (2021). Big Data and Predictive Analytics for Customer Retention: Exploring the Role of Machine Learning in E-Commerce. International Journal of Emerging Trends in Computer Science and Information Technology, 2(2), 26-34.

[38] Penmetsa, M., Bhumireddy, J. R., Chalasani, R., Tyagadurgam, M. S. V., Gangineni, V. N., & Pabbineedi, S. (2021). Next-Generation Cybersecurity: The Role of AI and Quantum Computing in Threat Detection. International Journal of Emerging Trends in Computer Science and Information Technology, 2(4), 54-61.

[39] Polu, A. R., Vattikonda, N., Gupta, A., Patchipulusu, H., Buddula, D. V. K. R., & Narra, B. (2021). Enhancing Marketing Analytics in Online Retailing through Machine Learning Classification Techniques. Available at SSRN 5297803.

[40] Polu, A. R., Buddula, D. V. K. R., Narra, B., Gupta, A., Vattikonda, N., & Patchipulusu, H. (2021). Evolution of AI in Software Development and Cybersecurity: Unifying Automation, Innovation, and Protection in the Digital Age. Available at SSRN 5266517.

[41] Polu, A. R., Vattikonda, N., Buddula, D. V. K. R., Narra, B., Patchipulusu, H., & Gupta, A. (2021). Integrating AI-Based Sentiment Analysis with Social Media Data for Enhanced Marketing Insights. Available at SSRN 5266555.

[42] Buddula, D. V. K. R., Patchipulusu, H. H. S., Polu, A. R., Vattikonda, N., & Gupta, A. K. (2021). INTEGRATING AI-BASED SENTIMENT ANALYSIS WITH SOCIAL MEDIA DATA FOR ENHANCED MARKETING INSIGHTS. Journal Homepage: http://www. ijesm. co. in, 10(2).

[43] Gupta, A. K., Buddula, D. V. K. R., Patchipulusu, H. H. S., Polu, A. R., Narra, B., & Vattikonda, N. (2021). An Analysis of Crime Prediction and Classification Using Data Mining Techniques.

[44] Rajiv, C., Mukund Sai, V. T., Venkataswamy Naidu, G., Sriram, P., & Mitra, P. (2022). Leveraging Big Datasets for Machine Learning-Based Anomaly Detection in Cybersecurity Network Traffic. J Contemp Edu Theo Artific Intel: JCETAI/102.

[45] Sandeep Kumar, C., Srikanth Reddy, V., Ram Mohan, P., Bhavana, K., & Ajay Babu, K. (2022). Efficient Machine Learning Approaches for Intrusion Identification of DDoS Attacks in Cloud Networks. J Contemp Edu Theo Artific Intel: JCETAI/101.

[46] Bhumireddy, J. R., Chalasani, R., Tyagadurgam, M. S. V., Gangineni, V. N., Pabbineedi, S., & Penmetsa, M. (2020). Big Data-Driven Time Series Forecasting for Financial Market Prediction: Deep Learning Models. Journal of Artificial Intelligence and Big Data, 2(1), 153–164.DOI: 10.31586/jaibd.2022.1341

[47] Nandiraju, S. K. K., Chundru, S. K., Vangala, S. R., Polam, R. M., Kamarthapu, B., & Kakani, A. B. (2022). Advance of AI-Based Predictive Models for Diagnosis of Alzheimer's Disease (AD) in healthcare. Journal of Artificial Intelligence and Big Data, 2(1), 141–152.DOI: 10.31586/jaibd.2022.1340

[48] Tyagadurgam, M. S. V., Gangineni, V. N., Pabbineedi, S., Penmetsa, M., Bhumireddy, J. R., & Chalasani, R. (2022). Designing an Intelligent Cybersecurity Intrusion Identify Framework Using Advanced Machine Learning Models in Cloud Computing. Universal Library of Engineering Technology, (Issue).

[49] Vangala, S. R., Polam, R. M., Kamarthapu, B., Kakani, A. B., Nandiraju, S. K. K., & Chundru, S. K. (2022). Leveraging Artificial Intelligence Algorithms for Risk Prediction in Life Insurance Service Industry. Available at SSRN 5459694.

[50] Polam, R. M., Kamarthapu, B., Kakani, A. B., Nandiraju, S. K. K., Chundru, S. K., & Vangala, S. R. (2021). Data Security in Cloud Computing: Encryption, Zero Trust, and Homomorphic Encryption. International Journal of Emerging Trends in Computer Science and Information Technology, 2(3), 70-80.

[51] Gangineni, V. N., Pabbineedi, S., Penmetsa, M., Bhumireddy, J. R., Chalasani, R., & Tyagadurgam, M. S. V. Efficient Framework for Forecasting Auto Insurance Claims Utilizing Machine Learning Based Data-Driven Methodologies. International Research Journal of Economics and Management Studies IRJEMS, 1(2).

[52] Vattikonda, N., Gupta, A. K., Polu, A. R., Narra, B., Buddula, D. V. K. R., & Patchipulusu, H. H. S. (2022). Blockchain Technology in Supply Chain and Logistics: A Comprehensive Review of Applications, Challenges, and Innovations. International Journal of Emerging Research in Engineering and Technology, 3(3), 99-107.

[53] Narra, B., Vattikonda, N., Gupta, A. K., Buddula, D. V. K. R., Patchipulusu, H. H. S., & Polu, A. R. (2022). Revolutionizing Marketing Analytics: A Data-Driven Machine Learning Framework for Churn Prediction. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 3(2), 112-121.

[54] Polu, A. R., Narra, B., Buddula, D. V. K. R., Patchipulusu, H. H. S., Vattikonda, N., & Gupta, A. K. BLOCKCHAIN TECHNOLOGY AS A TOOL FOR CYBERSECURITY: STRENGTHS, WEAKNESSES, AND POTENTIAL APPLICATIONS.

[55] Bhumireddy, J. R., Chalasani, R., Tyagadurgam, M. S. V., Gangineni, V. N., Pabbineedi, S., & Penmetsa, M. (2022). Big Data-Driven Time Series Forecasting for Financial Market Prediction: Deep Learning Models. Journal of Artificial Intelligence and Big Data, 2(1), 153–164.DOI: 10.31586/jaibd.2022.1341

[56] Nandiraju, S. K. K., Chundru, S. K., Vangala, S. R., Polam, R. M., Kamarthapu, B., & Kakani, A. B. (2022). Advance of AI-Based Predictive Models for Diagnosis of Alzheimer's Disease (AD) in healthcare. Journal of Artificial Intelligence and Big Data, 2(1), 141–152.DOI: 10.31586/jaibd.2022.1340

[57] Pabbineedi, S., Kakani, A. B., Nandiraju, S. K. K., Chundru, S. K., Tyagadurgam, M. S. V., & Gangineni, V. N. (2023). Scalable Deep Learning Algorithms with Big Data for Predictive Maintenance in Industrial IoT. International Journal of AI, BigData, Computational and Management Studies, 4(1), 88-97.

[58] Chalasani, R., Vangala, S. R., Polam, R. M., Kamarthapu, B., Penmetsa, M., & Bhumireddy, J. R. (2023). Detecting Network Intrusions Using Big Data-Driven Artificial Intelligence Techniques in Cybersecurity. International Journal of AI, BigData, Computational and Management Studies, 4(3), 50-60.

[59] Vangala, S. R., Polam, R. M., Kamarthapu, B., Penmetsa, M., Bhumireddy, J. R., & Chalasani, R. (2023). A Review of Machine Learning Techniques for Financial Stress Testing: Emerging Trends, Tools, and Challenges. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 4(1), 40-50.

[60] Kakani, A. B., Nandiraju, S. K. K., Chundru, S. K., Tyagadurgam, M. S. V., Gangineni, V. N., & Pabbineedi, S. (2023). A Survey on Regulatory Compliance and AI-Based Risk Management in Financial Services. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 4(4), 46-53.

[61] Bhumireddy, J. R., Chalasani, R., Vangala, S. R., Kamarthapu, B., Polam, R. M., & Penmetsa, M. (2023). Predictive Machine Learning Models for Financial Fraud Detection Leveraging Big Data Analysis. International Journal of Emerging Trends in Computer Science and Information Technology, 4(1), 34-43.

[62] Gangineni, V. N., Pabbineedi, S., Kakani, A. B., Nandiraju, S. K. K., Chundru, S. K., & Tyagadurgam, M. S. V. (2023). AI-Enabled Big Data Analytics for Climate Change Prediction and Environmental Monitoring. International Journal of Emerging Trends in Computer Science and Information Technology, 4(3), 71-79.

[63] Polam, R. M. (2023). Predictive Machine Learning Strategies and Clinical Diagnosis for Prognosis in Healthcare: Insights from MIMIC-III Dataset. Available at SSRN 5495028.

[64] Narra, B., Gupta, A., Polu, A. R., Vattikonda, N., Buddula, D. V. K. R., & Patchipulusu, H. (2023). Predictive Analytics in E-Commerce: Effective Business Analysis through Machine Learning. Available at SSRN 5315532.

[65] Narra, B., Buddula, D. V. K. R., Patchipulusu, H. H. S., Polu, A. R., Vattikonda, N., & Gupta, A. K. (2023). Advanced Edge Computing Frameworks for Optimizing Data Processing and Latency in IoT Networks. JOETSR-Journal of Emerging Trends in Scientific Research, 1(1).

[66] Patchipulusu, H. H. S., Vattikonda, N., Gupta, A. K., Polu, A. R., Narra, B., & Buddula, D. V. K. R. (2023). Opportunities and Limitations of Using Artificial Intelligence to Personalize E-Learning Platforms. International Journal of AI, BigData, Computational and Management Studies, 4(1), 128-136.

[67] Madhura, R., Krishnappa, K. H., Shashidhar, R., Shwetha, G., Yashaswini, K. P., & Sandya, G. R. (2023, December). UVM Methodology for ARINC 429 Transceiver in Loop Back Mode. In 2023 3rd International Conference on Mobile Networks and Wireless Communications (ICMNWC) (pp. 1-7). IEEE.

[68] Shashidhar, R., Kadakol, P., Sreeniketh, D., Patil, P., Krishnappa, K. H., & Madhura, R. (2023, November). EEG data analysis for stress detection using k-nearest neighbor. In 2023 International Conference on Integrated Intelligence and Communication Systems (ICIICS) (pp. 1-7). IEEE.

[69] KRISHNAPPA, K. H., & Trivedi, S. K. (2023). Efficient and Accurate Estimation of Pharmacokinetic Maps from DCE-MRI using Extended Tofts Model in Frequency Domain.

[70] Krishnappa, K. H., Shashidhar, R., Shashank, M. P., & Roopa, M. (2023, November). Detecting Parkinson's disease with prediction: A novel SVM approach. In 2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE) (pp. 1-7). IEEE.

[71] Shashidhar, R., Balivada, D., Shalini, D. N., Krishnappa, K. H., & Roopa, M. (2023, November). Music Emotion Recognition using Convolutional Neural Networks for Regional Languages. In 2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE) (pp. 1-7). IEEE.

[72] Madhura, R., Krishnappa, K. H., Manasa, R., & Yashaswini, K. P. (2023, August). Slack Time Analysis for APB Timer Using Genus Synthesis Tool. In International Conference on ICT for Sustainable Development (pp. 207-217). Singapore: Springer Nature Singapore.

[73] Krishnappa, K. H., & Gowda, N. V. N. (2023, August). Dictionary-Based PLS Approach to Pharmacokinetic Mapping in DCE-MRI Using Tofts Model. In International Conference on ICT for Sustainable Development (pp. 219-226). Singapore: Springer Nature Singapore.

[74] Krishnappa, K. H., & Gowda, N. V. N. (2023, August). Dictionary-Based PLS Approach to Pharmacokinetic Mapping in DCE-MRI Using Tofts Model. In International Conference on ICT for Sustainable Development (pp. 219-226). Singapore: Springer Nature Singapore.

[75] Madhura, R., Krutthika Hirebasur Krishnappa. et al., (2023). Slack time analysis for APB timer using Genus's synthesis tool. 8th Edition ICT4SD International ICT Summit & Awards, Vol.3, 207–217. https://doi.org/10.1007/978-981-99-4932-8_20

[76] Shashidhar, R., Aditya, V., Srihari, S., Subhash, M. H., & Krishnappa, K. H. (2023). Empowering investors: Insights from sentiment analysis, FFT, and regression in Indian stock markets. 2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE), 01–06. https://doi.org/10.1109/AIKIIE60097.2023.10390502

[77] Jayakeshav Reddy Bhumireddy, Rajiv Chalasani, Mukund Sai Vikram Tyagadurgam, Venkataswamy Naidu Gangineni, Sriram Pabbineedi, Mitra Penmetsa. Predictive models for early detection of chronic diseases in elderly populations: A machine learning perspective. Int J Comput Artif Intell 2023;4(1):71-79. DOI: 10.33545/27076571.2023.v4.i1a.169