International Journal of Computational and Experimental Science and ENgineering
(IJCESEN)
Vol. 11-No.2 (2025) pp. 2383-2395
http://www.ijcesen.com
ISSN: 2149-9144

Research Article

A Novel Cognitive Multi-Label Classification Model for Social Media Data Based on Dolphin Optimized Learning and Hybrid Classification networks

Shumama Ansa[1*], G. Narsimha[2]

[1]Research Scholar, Computer Science and Engineering, Jawaharlal Nehru Technological University Hyderabad, Telangana, India
* **Corresponding Author Email:** shumama2011@gmail.com - **ORCID:** 0009-0007-7670-4185

[2]Professor and Principal, Computer Science and Engineering, JNTUH University College of Engineering Sultanpur, Telangana, India
Email: narsimha06@jntuh.ac.in - **ORCID:** 0000-0002-0143-8122

Article Info:

DOI: 10.22399/ijcesen.1737
Received : 08 February 2025
Accepted : 05 April 2025

Keywords :

Extreme Multi-Label Classification,
Dolphin Optimized Learning Model,
Stacked Gated Recurrent Units,
Feed Forward Networks.

Abstract:

Social media plays a pivotal role in people's daily lives where users distribute diverse materials and subjects like ideas, events, and emotions. As the number of people grows, extensive use of social platforms has resulted in the creation of vast amounts of information. These unstructured data need to be labelled for understanding the relevant information that aids for various applications such as healthcare, entertainment and even sentimental politics. These unstructured data have large number of labels which needs the brighter light of annotation that tags the document with the most relevant labels. Extreme Multi-Label Classification (XMTC) aims to solve the above problem by automatically labelling a file with the most pertinent label from the large buckets of the large label sets. Because of the surge in big data, implementing the XMTC has become significant challenge to handle the millions of data, features and labels. This bottleneck was overshadowed with the arrival of Machine Learning (ML) and Deep Learning (DL) algorithms. But the computational overhead in training these learning networks degrades the performance of XMTC for handling the larger social media data. To solve this aforementioned problem, this research paper proposes the ensemble combination of Dolphin Optimized Learning and Hybrid classification networks. The proposed model comprises of triple set: Initially, it incorporates the multi-label dolphin optimized learning procedure to recognize the weight of every word in relation to labels. The label structure and document details are utilized to ascertain the link among the phrases and labels to compress the labels. Finally, the label-aware classification networks formulated with the Stacked Gated Recurrent networks and Feed Forward networks to attain the final label-aware massive documents. The comprehensive experimentation is carried out using the EuroLeX benchmarks and various performance metrics like accuracy, precision, recall, hamming score are calculated. To prove the excellence of the recommended XMTC with the varied state-of-the art models, Results demonstrates the proposed model has exhibited the superior performance over the models notably on the tail labels.

1. Introduction

Extreme Multi-Label Text Classification (XMTC) focuses on effortlessly labeling a document with the pertinent tags from an immense label set. For example, there are lots of groups on Wikipedia, and one might want to develop a classifier that can note a given text with the set of most significant groups [1,2]. XMTC has gained increasing importance because of the explosion of big data, while it has become considerably more challenging since it must simultaneously manage vast amount of documents, features, and labels [3-5]. Therefore, it is crucial to create efficient XMTC for several real-world applications including product categorization in e-commerce, news tagging, and so on.

Unlike traditional multi-class classification, multi-label text Classification allows multiple labels to co-exist for a single record. In general, the numerous labels are often semantically related, and it is advantageous for the multi-label learning approach to utilize the correlation between distinct labels. At the same time, there might be a substantial count of tail labels with sparse supportive records, making them difficult to model. To address the above-mentioned challenges, researchers focus on dual aspects: (1) approach to illustrate labels so that the correlation between labels can be precisely mined, and (2) approach to demonstrate documents so that the interdependence between texts can be fully acquired.

1.1 Role of ML and DL Model in XMTC

In the recent times, DL models such as Convolutional Neural Networks (CNN) [6], Classification Neural Networks (CLNN) [7], Recurrent Neural networks (RNN) [8] have attained the great success to represent the text data. But these techniques are limited to the same level of representation of text data. Besides the success of these deep learning models, Attentive Neural Networks (ANN), Attention XML [9] and EXAM [10] has also gained the more popularity in solving the XMTC problems. But these algorithms concentrate solely on the document text but disregard the label arrangement within extreme labels, that is demonstrated to be crucial in XMTC learning problems. Additionally, computational complexity fuels the existing model to achieve the inaccurate classification.

1.2 Motivation of the Research

Existing Deep learning Models has exhibited the inaccurate classification due to ignoring the label structures and suffers from the computational complexity which persists to be real problem among the researchers. To solve this aforementioned problem, this research paper proposes the hybrid combination of Dolphin Optimized Neural Networks (DONN) and Stacked GRUs with the Feed Forward networks to achieve the better XMTC for the massive social media text data. The proposed framework consists of three components such as Label Contributor Phase (LCP), Compression Phase (CP) and Label Aware Classification Networks (LACN).

1.3 Contribution of the Research Article

a) The proposed paper proposes the Dolphin Optimized Deep Neural Networks for detecting the relationship of the words to the label. The DONN models with self –attention maps construct the label-aware document illustration by concurrently examining the content and label structure.

b) The Hybrid Stacked GRU networks are proposed for achieving the better classification using the larger labels and documents.

c) The Label aware semantic relationship using attention maps has been proposed to extract the relationship between every document and all the multiple labels for XMTC.

d) Extensive Experimentation has been carried out using the EURLlEX57K datasets in which the performance metrics are measured and evaluated against other state-of-the-art procedures. Experimental findings highlighted the recommended approach has exhibited superior XMTC performance.

1.4 Structure of the Paper

The sequential arrangement of the manuscript is presented pursues:1) Section-2 demonstrates the relevant studies by more than different authors. The proposed system, data-preprocessing, classifier architectures are presented in Section-3. The experimental details, comparative analysis and results discussion are discussed in Section-4. At last, the study concludes with the future enhancement in Section-5.

2. Related Works

Bayu and Tegegn (2024) [11] conducted pioneering research on multi-label emotion classification for Amharic social media text, addressing a critical gap in the field. They collected 18,000 datasets from social media platforms annotated by psychologists and professionals. Utilizing word2vec and one-hot encoding, they trained four DL models: LSTM, BiLSTM, CNN, and GRU. The best accuracy was achieved with BiLSTM at 54.5%, followed by CNN at 54%, and LSTM at 53.1%. However, the study's primary drawback lies in its relatively low overall accuracy, with GRU performing particularly poorly at 39.7%. The limited dataset size and the complexity of Amharic language processing pose significant challenges for comprehensive emotion classification. Wang et al. (2024) [12] introduced a novel deep Active Learning method relies on Bayesian DL and Expected confidence (BEAL) for multi-label text classification, addressing the challenge of limited labeled data. Their approach leverages Bayesian deep learning to obtain the method's subsequent forecast probability distribution and specifies an innovative expected confidence-based acquisition function. Experimental results with a BERT-based multi-label classification (MLTC) technique demonstrated more efficient model training, enabling convergence with less labeled samples. But, the computational complexity of Bayesian deep learning approaches may pose challenges for real-world implementation. Samy (2023) [13] explored sentiment analysis using hybrid BERT models to enhance classification accuracy across different scenarios. The research developed four deep learning models by combining BERT with BiLSTM and Bidirectional GRU algorithms. They utilized pre-trained word embedding vectors to assist the algorithm fine-tuning process. The developed models aimed to improve accuracy and evaluate the impact of hybridizing BiGRU and BiLSTM layers. The architectures incorporating BiGRU layers achieved the best results. However, the computational intensity of the hybrid models and their demand for extensive computational resources remain significant drawbacks. De and Vats (2023) [14] developed a robust multi-label classifier to classify diverse concerns expressed in tweets about vaccination. Their comprehensive approach utilized advanced natural language processing methods and ML models, such as transformer models like BERT and GPT 3.5, alongside traditional methods such as Classifier Chains, Support Vector Machine (SVM), Random Forest (RF), and Naive Bayes (NB). The cutting-edge large language models outperformed other methods, demonstrating the potential of advanced AI in understanding complex social media discourse. Although, key drawbacks include potential bias in data collection and difficulty capturing subtle vaccine-related sentiments. Ameer et al. (2023) [15] advanced multi-label emotion classification by introducing innovative transfer learning techniques with multiple attention mechanisms. Their approach utilized transformer networks like XLNet, DistilBERT, and RoBERTa to reveal each word's contribution to emotion classification. The XLNet-multi-attention model showed 45.6% accuracy on the Ren-CECps Chinese dataset. Although the results are impressive, the model's complexity, high computational demands, and difficulties in handling different languages and emotions limit its effectiveness. Paranjape et al. (2023) [16] evaluated several transformer-driven approaches by fine-tuning them for multi-label classification. Oversampling techniques were applied to address the class imbalance in the dataset. Ensemble methods were employed to enhance the system's effectiveness. The technique's reliance on multiple sophisticated models may limit its practicality for resource-constrained environments. Elangovan and Sasikala (2022) [17] introduced a novel approach is suggested utilizing the Enhanced Embedding from Language Model (EnELMo) for categorizing tweets into various classes. The proposed EWECNN approach consists of an elevated ELMo component to process Crisis word vectors. Their suggested method surpasses the classification of microblog texts with an improved accuracy rate outperforming other techniques. However, significant drawbacks include the limited generalizability to other domains, and the high computational requirements of the complex model architecture. Ashraf et al. (2022) [18] created the first multi-label emotion dataset for Urdu, comprising 6,043 tweets representing six basic emotions. Recognizing the morphological and syntactic challenges of the Urdu language, they developed baseline classifiers, including ML, DL approaches and BERT. The study employed various text representations, like stylometric attributes, pre-trained word embeddings, and n-grams. Nevertheless, the main drawback was the dataset's small size. Khalil et al. (2021) [19] proposed a multi-label emotion classification model for Arabic tweets utilising BiLSTM deep network. Their approach included comprehensive preprocessing steps, such as Arabic language stemming, emoji text replacement. The study uniquely examined the effect of hyperparameter tuning on model performance, achieving a 9% improvement in validation accuracy. Nonetheless, the method's effectiveness may be constrained by its reliance on Aravec embeddings, which could limit adaptability to evolving language usage.

Bdeir and Ibrahim (2020) [20] developed an architecture for Arabic tweet multi-label classification using word embedding techniques and deep learning models. They constructed a dataset of 160,000 Arabic tweets and compared two DL methods:

CNN and RNN. Their outcomes demonstrated nearly identical performance across both network types, with accuracy scores around 90% and Hamming loss approximately 0.02.

However, the study's reliance on a single dataset and the lack of evaluation across diverse linguistic contexts limit its generalizability.

3. Proposed Methodology

3.1 System Overview

Figure 1 presents the proposed framework for an efficient XMTC model. As shown in Figure 1, proposed XMTC framework consists for four components such as Data Pre-processing, DONN Models for Handling the Multi-label documents, Feature extraction and finally classification layers. The detailed description of each and every component is as follows

Table 1. *Quick Summary of the Different Related Works*

S. No	Authors	Year	Technology	Results Obtained	Advantages	Drawback
1	Bayu and Tegegn	2024	Word2vec, One-hot encoding, Deep Learning (LSTM, BiLSTM, CNN, GRU)	BiLSTM: 54.5% accuracy, CNN: 54%, LSTM: 53.1%	Utilized multiple DL approaches	Low overall accuracy, Small dataset size, Complex language processing
2	Wang et al.	2024	Bayesian Deep Learning, BEAL, BERT-based MLTC	Efficient model training with fewer labeled samples	Innovative expected confidence-based acquisition function	High computational complexity of Bayesian deep learning
3	Samy	2023	Hybrid BERT with BiLSTM and BiGRU, Pre-trained word embedding	Best results with BiGRU layer architectures	Improved accuracy through model hybridization	High computational resource demands
4	De and Vats	2023	Transformer models (BERT, GPT 3.5), Traditional ML (SVM, RF, NB)	Large language models outperformed traditional methods	Comprehensive approach to understanding social media discourse	Potential data collection bias, Difficulty capturing subtle sentiments
5	Ameer et al.	2023	Transformer networks (XLNet, DistilBERT, RoBERTa) with Multi-attention	RoBERTa-Multi-attention: 62.4% accuracy (SemEval-2018), XLNet-Multi-attention: 45.6% accuracy (Ren-CECps)	Innovative transfer learning with attention mechanisms	Complex model, High computational demands
6	Paranjape et al.	2023	Transformer models (XLNet, Longformer, BERT, BigBird)	Macro F1-score: 0.5649 (official phase), 0.6605 (post-competition)	Oversampling and ensembling techniques	Reliance on multiple sophisticated models
7	Elangovan and Sasikala	2022	Improved ELMo, CNN-RNN approaches	93.46% accuracy, 92.99% F1-Score	High performance in microblog text classification	Limited generalizability, High computational requirements
8	Ashraf et al.	2022	ML, DL and BERT-based classifiers	Comprehensive evaluation using multiple metrics	First multi-label emotion dataset for Urdu	Small dataset size
9	Khalil et al.	2021	BiLSTM with Aravec word embedding	Improved validation accuracy, Higher micro F1 score	Comprehensive preprocessing, Hyperparameter tuning	Limited adaptability to evolving language
10	Bdeir and Ibrahim	2020	CNN and RNN with word embedding	~90% accuracy, ~0.02 Hamming loss	Extensive dataset of 160,000 tweets	Limited generalizability, Single dataset evaluation

3.2 Data-Preprocessing technique

Before training the proposed model with the datasets, mentioned in [21], pre-processing technique is adopted for the parallel datasets by adopting the following steps:

(i) Converting all the texts into lower case
(ii) Excluding special characters from text, except apostrophes
(iii) tokenizing the source and target parallel sentences into sub word tokens using Keras Libraries.
(iv) generating the sub word embeddings as input to the DONN models.

3.3 DONN Models for Handling the Multiple Labels

This section discusses about the general overview of the dolphin optimization model, Dolphin optimized model and label aware documents.

Dolphin Optimization Model

In this paper, a new optimization technique inspired by dolphin echolocation is introduced. This approach simulates the hunting strategies dolphins employ, where they use sonar to detect and adjust to the location of their prey. This process of modifying sonar signals to pinpoint the target is emulated as the core feature of the proposed method. Figure 2 is living dolphin grabbing its meal.

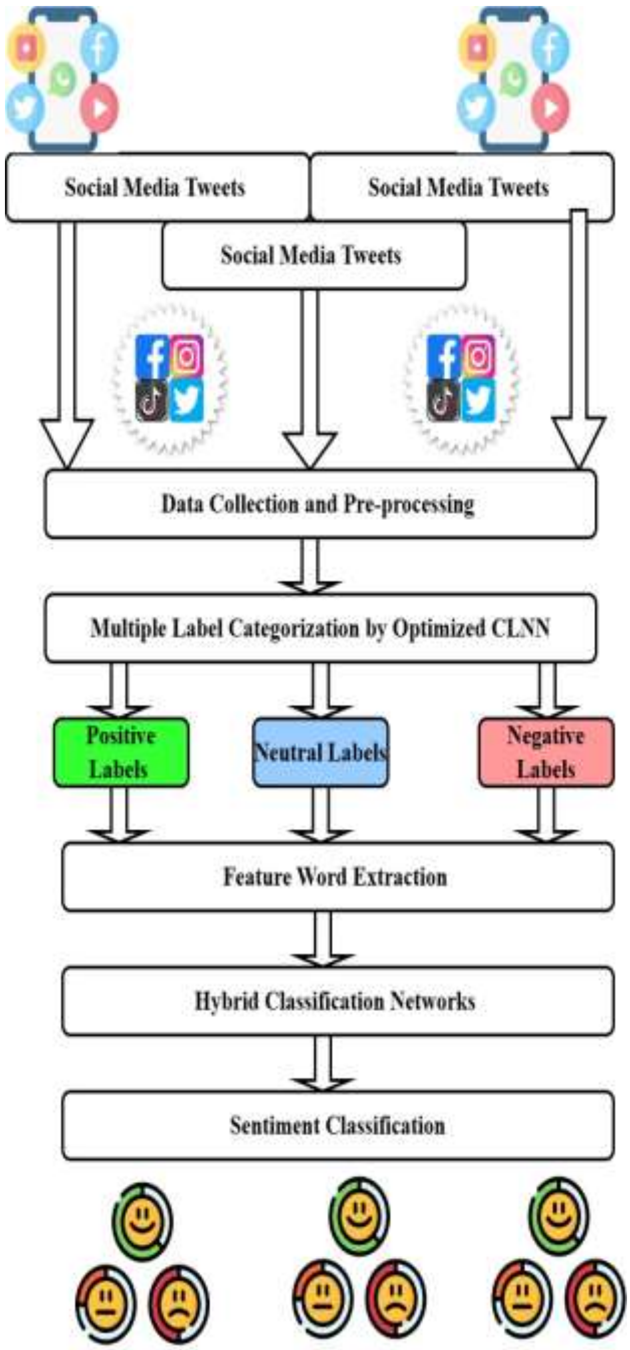

Figure 1. *Proposed XMTC Framework for the Multi-Label Classification*

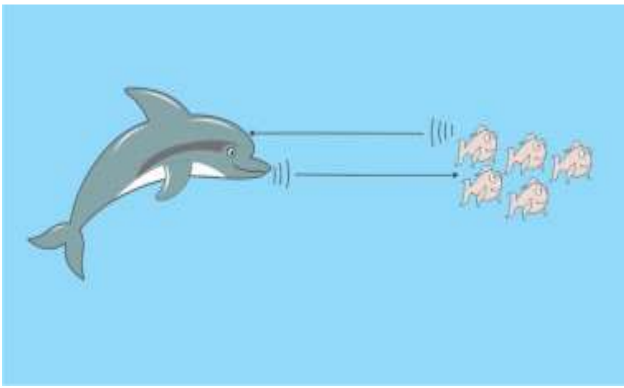

Figure 2. *A living dolphin grabbing its meal.*

Dolphin echolocation in nature

The concept of "echolocation" was introduced by Griffin [22,23] to detail the capability of flying bats to detect obstacles and prey by heeding to the echoes reflected from the large-frequency clicks they produce. Echolocation is employed by several mammals and a small number of bird species. The most well-researched form of echolocation is found in marine mammals, particularly in bottlenose dolphins. A dolphin generates sound waves in the form of clicks, with frequencies that are higher than those utilized for communication and that vary among species. When these sounds hit an item, few sound energy is reflected back towards the dolphin. Upon receiving the echo, it emits another click. The time interval among the emission of the click and the return of the echo helps the dolphin assess the distance to the object. The dolphin can also gauge the direction of the object based on the varying intensity of the signal retrieved on each side of its head. By repeatedly radiating clicks and retrieving the echoes, it is able to track and approach objects. Bats utilize echolocation, their sonar systems vary than those of dolphins. Bats generally rely on their sonar within shorter ranges, about 3–4 meters, they can recognize targets from a range of 10-100 meters. Bats typically track for fast-moving insects, which differs significantly than the traits of fish pursued by dolphins. Furthermore, Sound travels in air at roughly one-fifth the velocity it does in water, meaning that the rate of information exchange for bats at sonar transmission is minimal than that of dolphins. These environmental and prey-related differentiations necessitate distinct sonar systems, making a direct comparison between the two animals difficult

Dolphin echolocation (DE) optimization

The concept of DE can be compared to optimization in several ways. In the process of finding prey using echolocation, dolphins explore their surroundings before honing in on the target, which resembles the process of attaining the ideal solution to a problem. Initially, dolphins search the environment broadly, but as they approach their prey, they restrict their search and increase the frequency of their echolocation clicks to zero in on the target. This process can be modeled in optimization problems by limiting the exploration phase based on the proximity to the target. The optimization process can be categorized into two phases: in the first stage, the algorithm performs a global search by exploring various points in the solution space, seeking unexplored areas. In the second phase, the algorithm focuses more locally around the better solutions found during the first phase. This behavior is common to many metaheuristic algorithms. It encompasses modifying the proportion of solutions produced in the global search phase (phase 1) compared to those produced in the local search phase (phase 2). By utilizing the DE approach, this ratio can be modified according to a predefined curve. This curve dictates how the approach gradually shifts from global search to local search.The user details a curve that specifies how the optimization should converge. The algorithm then adjusts its parameters to follow this curve. In essence, the method considers the probability of selecting the best solution found so far compared to other alternatives in each iteration. This approach reduces the dependence on the parameters, as the curve defines the convergence criterion. The curve can be any smooth, increasing function, with some recommendations on its form, which will be further discussed. Previous studies have shown a unified approach to parameter selection in metaheuristics [24]. In these methods, a metric known as the Convergence Factor (CF) is used. The CF represents the average likelihood of selecting the best solution. For instance, in an example where steel profiles are assigned to a structure with four elements, the CF is calculated as the mean of the frequencies of each modal profile for those elements. This method allows for consistent convergence tracking in optimization problems. Before proceeding with the optimization process, the search space must be arranged using the following criterion:

Search Space Arrangement

For every variable being optimized at the procedure, sort the available options in either ascending or descending order. In cases where the options involve multiple characteristics, the sorting should prioritize the most significant one. As a result, for a given variable *j*, length *LAj* is formed, containing all the potential variants for the variable *j*. Then arranged side by side to form a matrix, *Matrix Alternatives MA*NV*, where *MA* is the maximum of *LAj* for each *j = 1:NV*, with *NV* depicts the total variables count. It is necessary to assign a curve that dictates how the convergence factor (CF) should evolve at the optimization phase. The variation in CF follows the designated curve during the optimization procedure.

$$PP(Loop_i) = PP_1 + (1 - PP_1)\frac{Loop_i^{Power}-1}{(LoopsNumber)^{Power}-1} \quad (1)$$

The predefined probability is denoted as PP, with PP1 representing the CF during the initial loop, where the solutions are chosen at random. $Loop_i$ refers to the current iteration number, while Power signifies the curve's degree. As illustrated, the curve in Equation (1) exhibits a Power degree. The loops count, also referred to as the Loop count, represents how many iterations the algorithm requires to reach the convergence point. This value must be selected by the user, depending on the computational resources available for the algorithm.

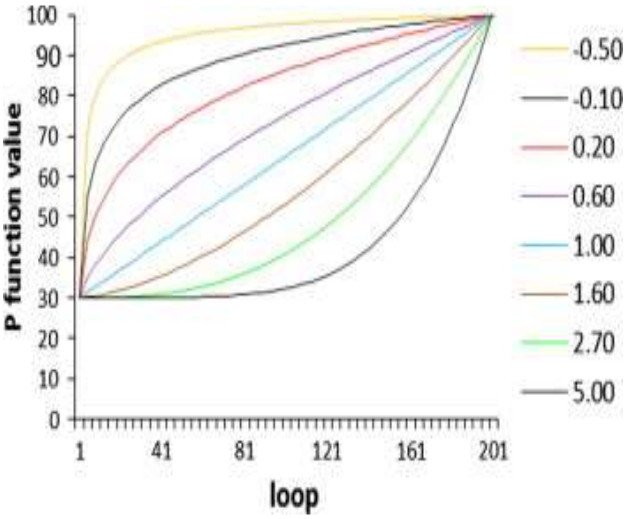

***Figure 3**. Overview Search Space Arrangement*

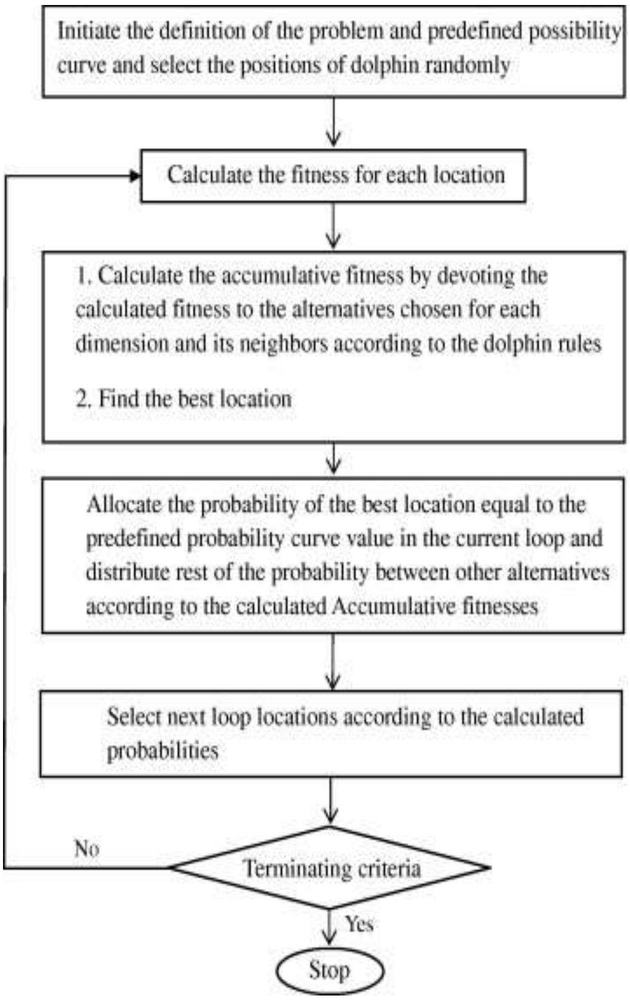

***Figure 4.** Working Flow of the Proposed Dolphin Optimization Model*

Figure 3 Alterations of *PP* by the moderations of the *Power*, utilizing the recommended formulae, Equation (1). Figure 4is working flow of the proposed dolphin optimization model.

Hyper-parameter Tuned Deep Neural Networks

Adjusting hyper-parameters (HP) is performed to acquire the optimal HP for the proposed framework to further lessen complexity. The HP that requires modification in this research wraps the count of hidden layers, number of hidden units, dropout rate, epochs, and batch size. The weights of the dense networks utilized by classification layers are fine-tuned by utilizing the straightforward DEO. The training network initially fetches the randomly chosen HP. The fitness value of the suggested approach is provided in equation (2). HP are calculated using Algorithm-1 during every iteration. The iteration concludes when the fitness function aligns with equation (2).

$$\boldsymbol{Fitness\ Function} = \{\boldsymbol{Max}(\boldsymbol{Accuracy}, \boldsymbol{Precision}, \boldsymbol{Recall}, \boldsymbol{F1-score}) \qquad \textbf{(2)}$$

The developed classification layer efficiently and with minimal computation identifies multiple labels of classes. Algorithm 1 outlines the proposed classification layers' functional procedure. According to the suggested algorithm, 78 hidden nodes, a momentum of 0.01, and 85 iterations are chosen.

Steps	Algorithm-1 // Pseudo Code for the Proposed Model
01	Input : Bias weight, concealed units, Epochs, Learning Rate
02	Target : Multi-Class Labels
03	Bias weight, concealed units, epochs, and learning rate should be assigned at random.
04	Set the three parameters
05	Utilize While loop for true
06	Utilize the formula (2) for determining the fitness function
07	Commence the For loop from t=1 to Max. iteration
08	Utilize equations (1& 2) to allocate the bias weights & input layers
09	Utilize equation (2) for quantizing the fitness function
10	Check If condition for (Fitness function is equal to threshold)
11	jump to Step 14
12	Otherwise
13	jump to Step 06
14	Stop

3.4 DEO Tuned Model for Word and Feature Embedding

To Establish the recommended approach, pre-processed text data is then converted into the low-dimensional dense vectors by word embedding techniques. The extreme labels is encoded into dense vectors coexistence graph to facilitate the accurate identification of label correlation and local patterns. As discussed, better extract label information, label-coexist graph G= (V, S) from the training data. In this visual, V is formed by the label node set, with S signifying the edge set. An edge set is included. There will be an edge connecting the i^{th}label and the J^{th}document. For better extraction of label correlation information, label co-exist graph G=(V,S) was built using python libraries. The objective is to map the extreme labels into a low-dimensional latent space, ensuring that labels positioned closely within the graph share similar representations. To achieve this, the widely recognized and robust Node2Vec [25] technique is employed. This method effectively explores the diverse neighborhoods of labels using a flexible, biased random walk approach, incorporating both breadth-first and depth-first sampling strategies**.** At last every label will be categorize as the group of positive, neutral and negative. Implemented by a *r*-dimensional dense vector, i.e., $\mathbf{l}(i) \in \mathrm{R}r$ for the i-th label ($i = 1, \ldots, k$) and the whole label set could be presented by

$$\mathbf{L} = (\mathbf{l}(1), \mathbf{l}(2), \ldots, \mathbf{l}(k)) \in \mathrm{R}r{\times}k \,. \qquad (3)$$

The proposed model DEO seeks to enhance the representation of each document by leveraging both its content and the structure of the labels.

3.5 Self-Attention Maps

Recent studies predominantly focus on incorporating attention layers to reduce redundant features, thereby enhancing the accuracy of the classification process. The self-attention mechanism, also referred to as intra-attention, operates by generating three vectors—Q (Query), K (Key), and V (Value)—for each input sequence. Consequently, the input sequences from each layer are converted into output sequences. In essence, this technique aligns the Query with corresponding key-value pairs using scaled dot-product functions. In multilabel contexts, a document might fetch multiple labels, and it should reflect the context most strongly related to every label. To focus on the different words on the documents, self –attention is applied on the outputs of DEO evoked Neural networks.

Mathematically, dot product for self –attention is computed as pursues

$$F(K,Q) = Softmax((K), Q^T))/(V_K)^{\wedge}0.5 \tag{4}$$

Softmax is applied along the first dimension to illustrate the varying contributions of each word to each label.

3.6 Classification Layers

For achieving the better classification for the multi-label data with the three classes, stacked GRU and feed forward networks are deployed. The detailed description of each and every module is presented as follows

Stacked GRU networks

The gating mechanism in GRU and LSTM RNNs mirrors the simple RNN in terms of parameterization. The weights associated with these gates are upgraded utilizing backpropagation through time (BTT) via stochastic gradient descent, aiming to diminish a loss or cost function [26]. Consequently, every parameter upgrade incorporates information related to the overall network state. The primary driving signal is the internal state of the network. Additionally, the adaptive parameter upgrades affect all aspects of the system's internal state [27]. This section examines three significant variants for deriving the gating equations, utilized consistently to both gates.

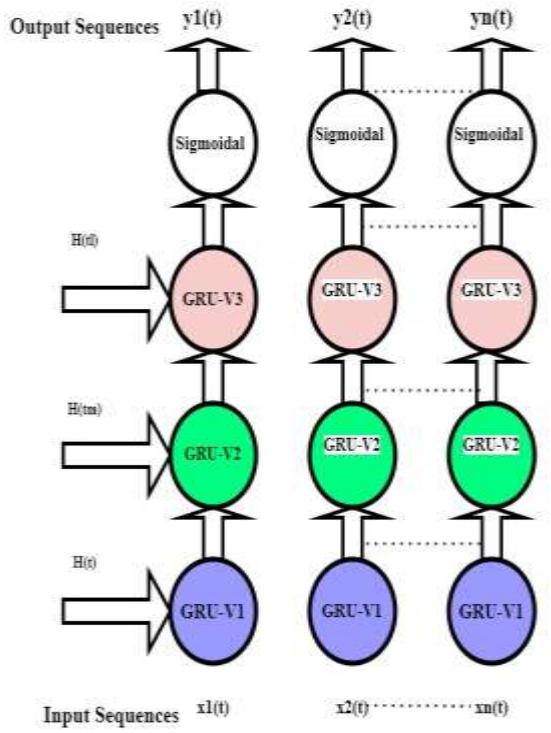

Figure 5. *Proposed Stacked GRU framework for deep feature extraction process*

GRU-First Variant

Here, every gate is calculated utilising only the previous hidden state and the bias, thereby decreasing the entire count of parameters by 2 × nm compared to the GRU RNN.

$$z_t = \sigma(U_z h_{t-1} + b_z) \tag{5}$$

$$r_t = \sigma(U_r h_{t-1} + b_r) \tag{6}$$

GRU-Second Variant

Every gate is calculated solely utilizing the prior hidden state, thereby decreasing the overall count of parameters by 2 × (nm + n) compared to the GRU RNN.

$$z_{tm} = \sigma(U_z h_{t-1}) \tag{7}$$

$$r_{tm} = \sigma(U_r h_{t-1}) \tag{8}$$

3.6.4 GRU-Third variant

Here, every gate is calculated solely utilising the bias, thereby diminishing the overall parameters count compared to the GRU RNN by 2 × (nm + n^2).

$$z_{tl} = \sigma(b_z) \tag{9}$$

$$r_{tl} = \sigma(b_r) \tag{10}$$

As mentioned in [28],GRU is a shallow model with weak capability of feature extraction, and the proposed GRU comprised of multiple variants of GRU units, as shown in Figure 5. The input of intial layer in GRU 's first variant is the original data and followed by the middle layer consist of the GRU-variant-2 structure. which is the output of the hidden layers of the upper layers composed of GRU-variant-3. Finally, the sigmoidal layers are added to the last layer of the GRU networks as mentioned in above equations.

The primary benefit of the recommended network lies in its potential for efficiently extracting high-level features from multi-label data. Furthermore, it effectively leverages information from time series, significantly enhancing classifier performance. Additionally, since the model parameters are not time-dependent, it eliminates the trade-off between time and accuracy highlighted in reference [29]. These low-level features are feed into the classification networks. These networks are designed based on the principles of Extreme Learning Machines (ELM). ELM represents a type of neural network characterized by a single hidden layer and operates on an auto-tuning principle. ELM demonstrates superior performance, faster execution, and reduced computational overhead compared to other learning techniques including SVM, Bayesian Classifiers (BC), K-Nearest

Neighbors (KNN), and even RF algorithm. A comprehensive explanation of the ELM's operational mechanism is provided in [30].
ELM is a kind of neural network that typically employs a single hidden layer, where the tuning of this layer is not necessarily essential. When compared to other models like Support Vector Machines (SVM) and RF, ELM demonstrates superior performance, faster processing speeds, and reduced computational demands. By leveraging kernel functions, ELM achieves high accuracy and efficient performance. The primary benefits of ELM include minimal training errors and enhanced approximation capabilities. ELM employs automated adjustment of weight biases and non-zero activation functions. The input feature mappings of ELM are represented by

$$Z = F(G(i), P) \tag{11}$$

The capsule features, characterized by dimension P, are represented as X. The function associated with the Extreme Learning Machine (ELM) output is described by

$$Y(i) = Z(i)\beta = Z(i)\, Z^T (\frac{1}{C} Z Z^T)^{-1} O \tag{12}$$

$$S = \alpha(\sum_{i=1}^{N}(Y(i), B(i), W(i)) \tag{13}$$

Let Z(i) represent the input feature maps. The temporal matrix β, is derived using the Moore-Penrose generalized inverse theorem, is represented as Z^T. Constants include C, while B and W denote the weights and biases of the network, respectively. The probability for every category is then determined using the softmax function

$$Y' = Softmax(S) \tag{14}$$

The predicted output Y′ aims to forecast the DFU mechanism using predefined datasets. To calculate the loss function, the cross-entropy function is employed, represented by the mathematical expression provided.

$$Loss = \left(\frac{1}{K}\right) \sum_{i=1}^{K} (Y(i) * Log\, Y' + \eta \left|\left|\theta\right|\right|^2 \tag{15}$$

In this context, K demonstrates the length of the feature vector for the dimensional capsule, η denotes the coefficient for regularization, and |θ| is the constant term.

4. Results and Discussions

In the experimentation, results, and analysis section of this proposed research focus shifts to a comprehensive examination of the empirical investigation conducted to assess the effectiveness of the algorithm presenting the most meticulous outcomes and comprehensive comparative analysis. Through the systematic exploration of the algorithm’s performance, effectiveness in multi-label classification in sentimental analysis is analysed.

4.1 Implementation Details

The proposed framework was deployed on a system equipped with the following hardware and software confirgation

a. **CPU and GPU:** an Intel I9 confirgation and NVIDIA Titan GPU was used for training the proposed framework. This combination provided the sufficient processing power for the deep learning models.
b. **Storage RAM:** The workstation has 16GB of RAM, which supports for efficient data handling and model training.
c. **Data Storage Space**: A total of 16 GB of diskspace are used for storing the datasets with the operating frequency of 3.4 GHZ.
d. **Software Design:** Python3.19 is used as the major programming tool for the implementing the proposed framework. Due to the versatility and its robustness, python is chosen for developing the framework. Several libraries such as Tensorflow, Keras, Matplotlib, Pandas, Numpy and Seaborn are utilised for developing the various tasks in the proposed framework.

4.2 Evaluation Metrics

These measures including accuracy, precision, recall, specificity, and F1-score are assessed and directly evaluated against varied state-of-the-art DL procedure used for multi-label classification to highlight the superiority of the recommended approach. Also, all these measures and statistical analyses are utilizing to highlight the higher proficiency of the recommended approach. Table 2 shows the mathematical expressions utilised for calculating these metrics. The early stopping method is applied to overcome generalization and overfitting concerns. This method halts the iteration when the validation performance of the proposed model no longer enhances over time. TP and TN represent True Positives and True Negatives, while FP and FN denote False Positives and False Negatives. First one is TP, where the values are correctly recognized as true and are actually true. Next is FP, where the values are incorrectly recognised as true when they are actually false.

Table 2. *Performance metrics used in the assessment*

Performance Metrics	Expression
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Recall	$\frac{\mathrm{TP}}{\mathrm{TP+FN}}$ x100
Specificity	$\frac{TN}{TN+FP}$
Precision	$\frac{TN}{TP+FP}$
F1-Score	$2.\frac{Precison * Recall}{Precision + Recall}$

Next one is FN, where the value is true yet is incorrectly recognised as negative. Last one is TN, where the value is negative and is correctly recognised as negative. Additionally, Normalized Discounted Cummulative Gain (NDCG) is calculated forcounting the relavent labels in the ground truth vector V. Higher the NDCG, is the higher the performance of the model

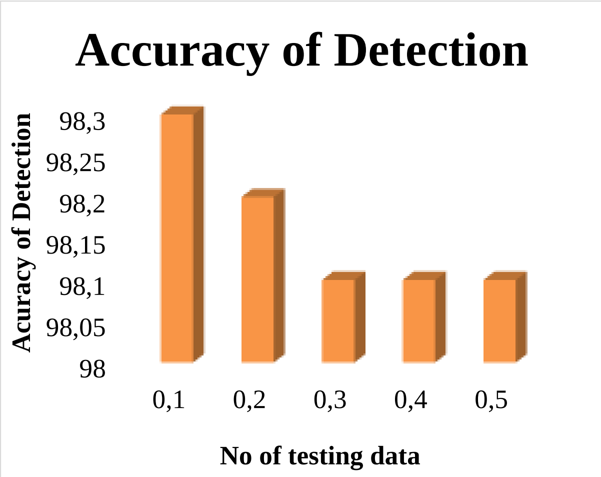

Figure 6. *Accuracy of Detection of the Proposed Model using EUROLEX57k Datasets*

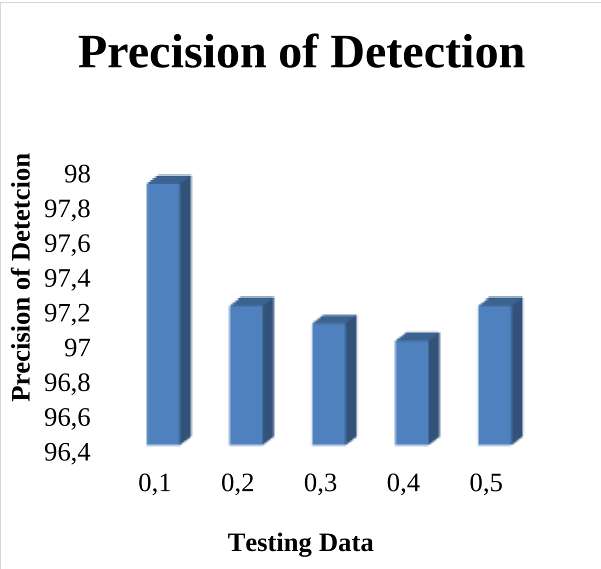

Figure 7. *Precision of Detection of the Proposed Model using EUROLEX57k Datasets*

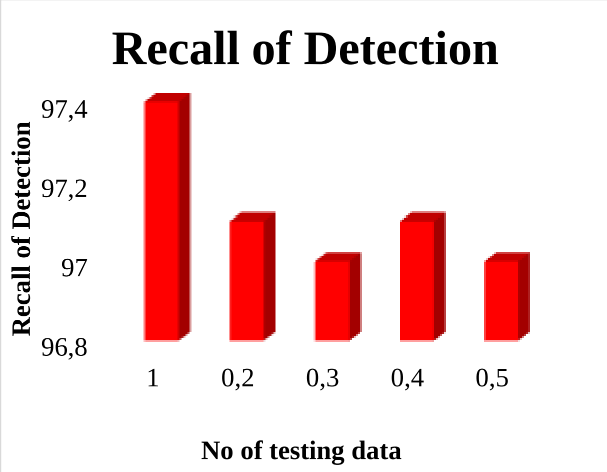

Figure 8. *Recall of Detection of the Proposed Model using EUROLEX57k Datasets*

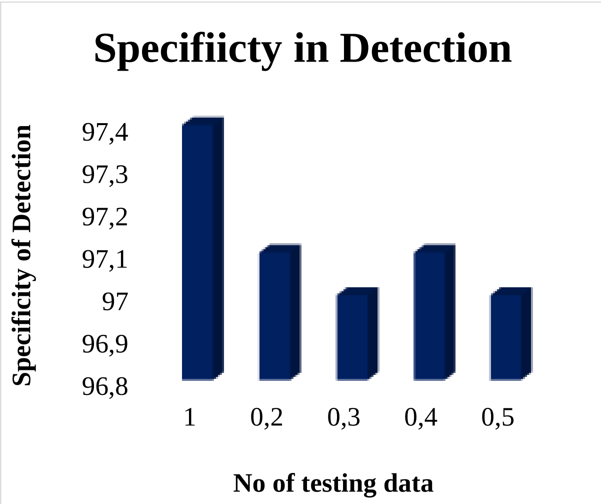

Figure 9. *Specificity of Detection of the Proposed Model using EUROLEX57k Datasets*

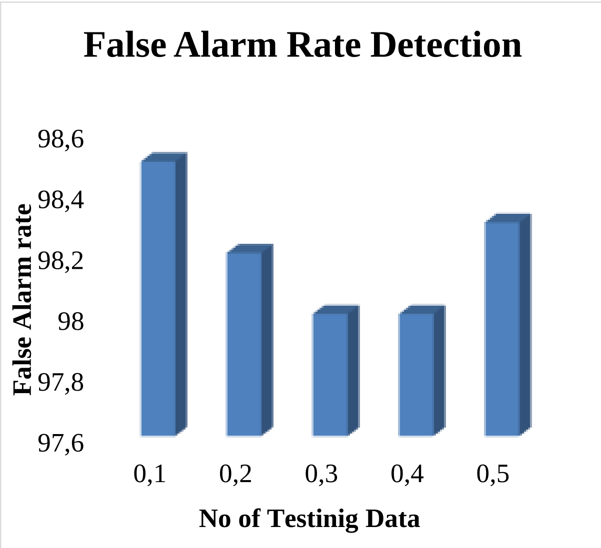

Figure 10*. False Alarm Rate Detection of the Proposed Model using EUROLEX57k Datasets*

Figure 6 presents the Accuracy of detection for the recommended approach in detecting the sentimental label. The model has produced the 98% of accuracy of detection in multi-label sentimental analysis. Figure 7-10 shows the precision, recall, specificity and F1-score of the recommended approach in detecting the multi-label sentiments. From the Figures, it is evident the effectiveness of the suggested method has given the stable performance in detecting sentimentals with the increase in testing word length. Figure 11 shows the NDCG performance of the algorithm in detecting the ground truth labels with the increase in the word-length. As the word length increases, NDCG performance is found to be 98.2% which is far better performance in multi-label classification.

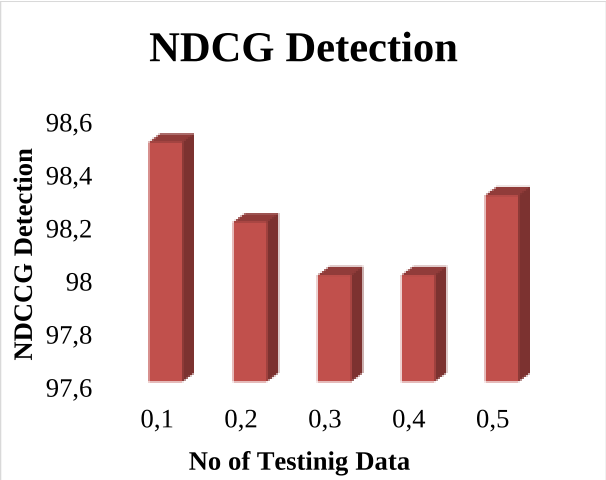

Figure 11. *NDCG Detection of the Proposed Model utilising EUROLEX57k Datasets*

4.3 Comparative Assessment

To prove the superiority of the recommended approach, models such as ATTENTIONXMTC, Non-Optimized GRU networks, LSTM networks, Bi-LSTM Models, Bi-LSTM+Attention models and Hybrid Attention Networks. Table 3 presents the comparative assessment of the metrics for various algorithms utilising EUROLEX57K datasets.Table 3 presents the evaluation metrics of distinct methods in classifying the multi-label sentimental documental data. From the table it is evident that suggested technique has produced superior results in multi-label classification and outperforms the other existing models in classifying the multi-label documents. Feed Forward Networks is applied to different fields [31-36].

5. Conclusion and Future Enhancement

In this article, novel XMTC framework was proposed which examines document content and label structure concurrently to obtain the discriminative content for every label meanwhile maintaining the label content. The proposed XMTC framework consist of ensemble of Dolphin Optimized neural network, Self –attention with the Stacked GRU for achieving the better performance in detecting multi-label classification. Extensive experiments are carried out using EUROLEX57k datasets and performance metrics like accuracy, precision, recall, specificity, F1-score, NSDC are measured and evaluated against other existing learning XMTC frameworks.

The performance of the recommended approach is found to have 98% of accuracy, 97.5% of precision, 97.2% of recall, 97% of F1-score and 98.4% of NSDC respectively. In a nutshell, the proposed model has provided the better platform to extract the better discriminative ability than baseline architectures. As future direction, this framework needs more brighter light in handling the real time massive multi-level datasets.

Table 3. *Performance Metrics of the Different Models in Classifying the Multi-labels using EUROLEX57K datasets.*

Metrics	Algorithms					
	A-XMLC	**GRU**	**Bi-LSTM**	**BiLA**	**HAN**	**Proposed Model**
Accuracy	90.0%	87.2%	915%	93.5%	94.6%	98.2%
Precision	89.2%	87.3%	90.2%	92.6%	93.4%	97.2%
Recall	88.7%	86.5%	90.1%	92.1%	93.0%	97.0%
Specificity	88.2%	87.4%	90.0%	90.5%	91.2%	98.2%
F1-Score	88.0%	87.6	89.4%	91.1%	93.5%	98.%
NDCG	**84.2%**	**83.8%**	**82.5%**	**87.4**	**89.4%**	**97.6%**

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.

- **Author contributions:** The authors declare that they have equal right on this paper.

- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1]Mishra, N. K., & Singh, P. K. (2022). Linear ordering problem-based classifier chain using genetic algorithm for multi-label classification. *Applied Soft Computing, 117*, 108395. https://doi.org/10.1016/j.asoc.2022.108395

[2]Zhao, D., Wang, Q., Zhang, J., & Bai, C. (2023). Mine diversified contents of multi-spectral cloud images along with geographical information for multi-label classification. *IEEE Transactions on Geoscience and Remote Sensing, 61*, 1–15.

[3]Liu, Z., Tang, C., Abhadiomhen, S. E., Shen, X. J., & Li, Y. (2023). Robust label and feature space co-learning for multi-label classification. *IEEE Transactions on Knowledge and Data Engineering*, 1–14. https://doi.org/ [Insert DOI if available]

[4]Singh, K., Sharma, B., Singh, J., Srivastava, G., Sharma, S., Aggarwal, A., & Cheng, X. (2020). Local statistics-based speckle reducing bilateral filter for medical ultrasound images. *Mobile Networks and Applications, 25*, 2367–2389.

[5]Huang, J., Qian, W., Vong, C. M., Ding, W., Shu, W., & Huang, Q. (2023). Multi-label feature selection via label enhancement and analytic hierarchy process. *IEEE Transactions on Emerging Topics in Computational Intelligence.*

[6]Koundal, D., Sharma, B., & Guo, Y. (2020). Intuitionistic based segmentation of thyroid nodules in ultrasound images. *Computers in Biology and Medicine, 121*, 103776.

[7]Kalpana, P., Malleboina, K., Nikhitha, M., Saikiran, P., & Kumar, S. N. (2024). Predicting cyberbullying on social media in the big data era using machine learning algorithm. *2024 International Conference on Data Science and Network Security (ICDSNS)*, Tiptur, India, 1–7. https://doi.org/10.1109/ICDSNS62112.2024.10691297

[8]Huang, J., Vong, C. M., Chen, C. P., & Zhou, Y. (2022). Accurate and efficient large-scale multi-label learning with reduced feature broad learning system using label correlation. *IEEE Transactions on Neural Networks and Learning Systems*, 1–14.

[9]Bayati, H., Dowlatshahi, M. B., & Hashemi, A. (2022). MSSL: A memetic-based sparse subspace learning algorithm for multi-label classification. *International Journal of Machine Learning and Cybernetics, 13*, 3607–3624.

[10]Zhu, X., Li, J., Ren, J., Wang, J., & Wang, G. (2023). Dynamic ensemble learning for multi-label classification. *Information Sciences, 623*, 94–111.

[11]Bayu, Yeshimebet & tegegn, tesfa. (2024). Multi-label Emotion Classification on Social Media Comments using Deep learning. 10.21203/rs.3.rs-4431629/v1.

[12]Wang, Q., Zhang, H., Zhang, W., & others. (2024). Deep active learning for multi-label text classification. *Scientific Reports, 14*, 28246. https://doi.org/10.1038/s41598-024-79249-7

[13]Samy, A. (2023). Sentiment analysis classification system using hybrid BERT models. *Journal of Big Data, 10*. https://doi.org/10.1186/s40537-023-00781-w

[14]De, S., & Vats, S. (2023). Decoding concerns: Multi-label classification of vaccine sentiments in social media. *ArXiv, abs/2312.10626*.

[15]Ameer, I., Bölücü, N., Siddiqui, M. H. F., Can, B., Sidorov, G., & Gelbukh, A. (2023). Multi-label emotion classification in texts using transfer learning. *Expert Systems with Applications, 213*(Part A), 118534. https://doi.org/10.1016/j.eswa.2022.118534

[16]Paranjape, A., Kolhatkar, G., Patwardhan, Y., Gokhale, O., & Dharmadhikari, S. (2023). *Converge at WASSA 2023 Empathy, Emotion and Personality Shared Task: A Transformer-based Approach for Multi-Label Emotion Classification*. Pune Institute of Computer Technology, Pune, India.

[17]Elangovan, A., & Sasikala, S. (2022). A multi-label classification of disaster-related tweets with enhanced word embedding ensemble convolutional neural network model. *Informatica: An International Journal, 46*(7). https://doi.org/10.31449/inf.v46i7.4280

[18]Ashraf, N., Khan, L., Butt, S., Chang, H.-T., Sidorov, G., & Gelbukh, A. (2022). Multi-label emotion classification of Urdu tweets. *PeerJ Computer Science*.

[19]Khalil, Enas & El Houby, Enas & Mohamed, Hoda. (2021). Deep Learning for emotion analysis in Arabic tweets. 10.21203/rs.3.rs-631537/v1.

[20]Bdeir, A. M., & Ibrahim, F. (2020). A framework for Arabic tweets multi-label classification using word embedding and neural networks algorithms. *In Proceedings of the 2020 2nd International Conference on Big Data Engineering (BDE '20)* (pp. 105–112). Association for Computing Machinery. https://doi.org/10.1145/3404512.3404526

[21]https://huggingface.co/datasets/NLP-AUEB/eurlex

[22]Kalpana, P., Kodati, S. S., Sreekanth, D., Smerat, N., & Akram, M. (2025). Explainable AI-driven gait analysis using wearable Internet of Things (WIoT) and human activity recognition. *Journal of Intelligent Systems and Internet of Things, 15*(2), 55–75. https://doi.org/10.54216/JISIoT.150205

[23]Dhiman, P., Kukreja, V., Manoharan, P., Kaur, A., Kamruzzaman, M. M., Dhaou, I. B., & Iwendi, C. (2022). A novel deep learning model for detection of severity level of the disease in citrus fruits. *Electronics, 11*(495).

[24]Kalpana, P., Narayana, P., Smitha, M. D., Keerthi, K., Smerat, A., & Akram, M. (2025). Health-Fots: A latency-aware fog-based IoT environment and efficient monitoring of body's vital parameters in

smart healthcare environment. *Journal of Intelligent Systems and Internet of Things, 15*(1), 144–156. https://doi.org/10.54216/JISIoT.150112

[25]Abirami, Abirami & Sanmugaraja, Lakshmanaprakash & R L, Priya & Hirlekar, Vaishali & Dalal, Bhargavi. (2024). Proactive Analysis and Detection of Cyber-attacks using Deep Learning Techniques. Indian Journal Of Science And Technology. 17. 1596-1605. 10.17485/IJST/v17i15.3044.

[26]Kalpana, P., Almusawi, M., Chanti, Y., Sunil Kumar, V., & Varaprasad Rao, M. (2024). A deep reinforcement learning-based task offloading framework for edge-cloud computing. *Proceedings of the 2024 International Conference on Integrated Circuits and Communication Systems (ICICACS)*, Raichur, India, 1–5. https://doi.org/10.1109/ICICACS60521.2024.10498232

[27]Salam, A., Ullah, F., Amin, F., & Abrar, M. (2023). Deep learning techniques for web-based attack detection in Industry 5.0: A novel approach. *Technologies, 11*(4), 107. https://doi.org/10.3390/technologies11040107

[28]Molaei, R., Rahsepar Fard, K., & Bouyer, A. (2025). A mathematical multi-objective optimization model and metaheuristic algorithm for effective advertising in the social Internet of Things. *Neural Computing and Applications, 37*, 4797–4821. https://doi.org/10.1007/s00521-024-10793-z

[29]Akhavan-Hejazi, Z. S., Esmaeili, M., & Ghobaei-Arani, M. (2024). Identifying influential users using a homophily-based approach in location-based social networks. *Journal of Supercomputing, 80*, 19091–19126.

[30]Saqia, B., Khan, K., Rahman, A. U., & others. (2025). Immoral post detection using a one-dimensional convolutional neural network-based LSTM network. *Multimedia Tools and Applications.* https://doi.org/10.1007/s11042-025-20757-7

[31]S, P., & A, P. (2024). Secured Fog-Body-Torrent : A Hybrid Symmetric Cryptography with Multi-layer Feed Forward Networks Tuned Chaotic Maps for Physiological Data Transmission in Fog-BAN Environment. *International Journal of Computational and Experimental Science and Engineering,* 10(4). https://doi.org/10.22399/ijcesen.490

[32]Hafez, I. Y., & El-Mageed, A. A. A. (2025). Enhancing Digital Finance Security: AI-Based Approaches for Credit Card and Cryptocurrency Fraud Detection. *International Journal of Applied Sciences and Radiation Research*, 2(1). https://doi.org/10.22399/ijasrar.21

[33]Fowowe, O. O., & Agboluaje, R. (2025). Leveraging Predictive Analytics for Customer Churn: A Cross-Industry Approach in the US Market. *International Journal of Applied Sciences and Radiation Researc*h 2(1). https://doi.org/10.22399/ijasrar.20

[34]Ibeh, C. V., & Adegbola, A. (2025). AI and Machine Learning for Sustainable Energy: Predictive Modelling, Optimization and Socioeconomic Impact In The USA. *International Journal of Applied Sciences and Radiation Research*, 2(1). https://doi.org/10.22399/ijasrar.19

[35]Sadula, V., & D. Ramesh. (2025). Enhancing Cross Language for English-Telugu pairs through the Modified Transformer Model based Neural Machine Translation. *International Journal of Computational and Experimental Science and Engineering*, 11(2). https://doi.org/10.22399/ijcesen.1740

[36]García, R., Carlos Garzon, & Juan Estrella. (2025). Generative Artificial Intelligence to Optimize Lifting Lugs: Weight Reduction and Sustainability in AISI 304 Steel. *International Journal of Applied Sciences and Radiation Researc*h, 2(1). https://doi.org/10.22399/ijasrar.22