EST: A Semi-supervised ensemble approach for Explainable AI-driven stress detection in social media

This Dissertation is Submitted in Fulfillment
of the Requirements for the Degree of

Bachelor of Science (B.Sc.)

in

Computer Science and Engineering (CSE)

by

Abdur Rahman (C191111)
Nurul Karim Symon (C183050)
Akther Uz Zaman (C183059)

TO
FACULTY OF SCIENCE AND ENGINEERING
INTERNATIONAL ISLAMIC UNIVERSITY CHITTAGONG
Autumn 22

DECLARATION

We hereby affirm the following statements regarding our thesis:

1. The thesis has been successfully completed as part of our undergraduate degree program at International Islamic University Chittagong.
2. The thesis work does not contain any previously published or third-party content without proper citation.
3. The thesis work has not been previously submitted for any other degree or diploma at any other university or institution.
4. We have appropriately acknowledged all significant sources of contribution in the thesis.

Student's Full Name and Metric ID:

Abdur Rahman	**C191111**
Nurul Karim Symon	**C183050**
Akther Uz Zaman	**C183059**

SUPERVISOR'S DECLARATION

I formally state that I have examined this thesis and claim it to be of sufficient quality and scope to be granted for the undergraduate degree of Bachelor of Science in Computer Science and Engineering.

Dr. Muhammed Jamshed Alam Patwary
Associate Professor
Department of Computer Science and Engineering
International Islamic University Chittagong

DEDICATION

This thesis report is dedicated to us, our supervisor and our family. The team work was satisfactory and the family's support was incredibly Amazing. Our dedicated and most hard working Supervisor who has been a constant support throughout these months. In this document, the contributions are acknowledged too.

ACKNOWLEDGMENT

To start with, All the praises to the Almighty Allah, for his mercy because of which we were able to finish our thesis despite having so many obstacles. Secondly, we would like to extend our gratitude to our supervisor, Dr. Muhammed Jamshed Alam Patwary for his continuous effort and guidelines from the very beginning of our research. And last but not least, without the prayers and well wishes of our parents, it would not have been possible.

ETHICAL STATEMENT

Hereby we state that, None of the unethical practices were used in the completion of our thesis work. The data we used for the research purpose are original. We carefully checked every citations we used here. The three writers of the work accept all the liabilities for any kind of violation of the thesis rule.

ABSTRACT

Social networking websites have become a vast 'archive' of human expression, containing human emotions from across the globe. Recognizing stress signals within this dynamic landscape is crucial for mental health monitoring, offering opportunities for timely intervention and support. This paper introduces a methodology integrating advanced machine learning techniques, notably Ensemble Learning and Self-Training, with Explainable Artificial Intelligence (XAI) to enhance stress detection capabilities. Our research follows a methodical approach, beginning with foundational steps of data collection and exploratory data analysis, providing insight into sentiments within social media interactions. Prioritizing the interpretability of model predictions, our methodology aims to build trust and offer meaningful insights for end-users and mental health professionals. Subsequent phases involve data preprocessing, and refining textual data to extract subtle indicators of stress. The core of our methodology is the fusion of Ensemble Learning and Self-Training, strategically combining diverse learners to iteratively refine the model using labeled and unlabeled data. With the help of XAI, we seek to deliver clearly and straightforwardly noticeable insights relating to decision-making produced by our stress detection model. Summing up the presented approach significantly contributes to mental health analytics taking a step forward toward better stress detection by a technical framework with an enhanced accuracy of 91.67%.

Keywords: Sentiments, Mental Health, Stress, Ensemble Learning, Self-Training, Explainable Artificial Intelligence (XAI).

Table of Contents

List of Figures

List of Tables

List of Algorithms

ABBREVIATION

The following list provides descriptions of various symbols and abbreviations that will be utilized in the subsequent sections of the document.

RF : Random Forest
DT : Decision Tree
KNN : K-Nearest Neighbors
NB : Naive Bayes
LR : Linear Regression
EST : EnsembleSelfTrain
SVM : Support Vector Machine
XAI : Explainable Artificial Intelligence
Bi-LSTM : Bidirectional Long Short-Term Memory

Chapter 1

Introduction

1.1 Overview

The widespread use of digital communication has turned social media platforms into vast spaces where people express their thoughts and feelings. In this extensive online world, detecting stress signals becomes crucial for monitoring mental health and providing timely support to those dealing with stress. Drawing inspiration from the works of Nijhawan et al. [1], KVTKN et al. [2], Khan and Sarkar [3], Yu and Sano [4] and others in the field, our thesis introduces a novel approach that integrates Explainable Artificial Intelligence (XAI) with Ensemble Learning and Self-Training to enhance stress detection in the domain of social media.

We started our exploration by taking essential steps in collecting and analyzing data and uncovering complex patterns in social media texts. Importantly, our method focuses on making model predictions easy to understand, aiming to build trust among users and mental health professionals. Following steps carefully prepare the data, distilling expressions related to stress. Building upon the insights provided by, Rabani et al. [5], Ansari et al. [6], Martino and Delmastro [7], Webb and Zheng [8] and others who explore the domain of ensemble learning, the fusion of Ensemble Learning and Self-Training culminates in a model that continuously refines itself through exposure to both labeled and unlabeled data.

Our thesis aims to provide insight into the XAI world regarding how one makes decisions using a stress detection model. By integrating XAI techniques, inspired by the works of Uban et al. [9], Angelov et al. [10], Arrieta et al. [11] and others, our objective is to provide transparent insights into the features and patterns influencing stress predictions, fostering a profound understanding of the model's outcomes.

In summary, "EST: A Semi-supervised ensemble approach for Explainable AI-driven stress detection in social media." represents noteworthy progress in developing strong and transparent stress detection models. With Ensemble Learning, Self-Training, and XAI at its core, this methodology seeks to provide effective and easily understandable solutions in the crucial domain of mental health monitoring.

This study presents several key contributions:

- Introduced a method of combining Ensemble Learning, Self-Training and XAI. Thus, increasing accuracy while maintaining transparency in the analytics of mental health.
- Enhanced stress detection by utilizing Word2Vec Word Embedding Models, enabling the model to decipher subtle stress patterns in diverse social media texts.
- Created a model which shows greater sensitivity to various stress indicators, contributing to a more resilient and effective stress signal decoding process.
- Ensured transparency and interpretability through the incorporation of Explainable AI techniques, allowing for deeper insight into what is happening in the stress detection algorithm.

1.2 Motivation and Scope of the Research

Influenced by increasing instances of stress in digital spaces, our research aims to develop a methodology at the intersection of machine learning and mental health analytics. The immediate need for early stress identification serves as the driving force, highlighting the significant contribution of cutting-edge machine learning. Our scope includes building a strong model using Explainable AI, Ensemble Learning, and Self-Training combined for improved transparency and accuracy. Our goal goes beyond theoretical developments to practical application, promoting mental health support in virtual spaces.

1.3 Problem Statement

In the era of widespread digital communication and social media interaction, there exists an escalating need for effective and interpretable methods to detect stress signals within the vast landscape of online expressions. The challenge lies in developing a stress detection model that not only achieves high accuracy but also provides transparent insights into its decision-making processes. Traditional machine learning models may lack the capacity to offer explanations for their predictions, hindering their utility in applications requiring a deep understanding of stress indicators. Addressing this gap becomes imperative for advancing mental health monitoring capabilities and facilitating timely support for individuals experiencing stressors in their online interactions. The research aims to devise an innovative solution by fusing Ensemble Learning and Self-Training methodologies. To clarify our results, we employed explainable AI, providing insights into how our model functions effectively. Additionally, we used visualizations to highlight specific words that significantly influence the results.

1.4 Research Objectives

The primary goals that we are tried to achieve with this work here:

- We aimed to develop and validate an integrated methodology (Ensemble Learning, Self-Training, and XAI) for stress detection in social media, with the goal of advancing accuracy and transparency in mental health analytics.
- Our objective was to investigate Word Embedding Models with Word2Vec to capture nuanced stress patterns. We seek to increase sensitivity to diverse indicators, making the stress detection model more robust and effective.
- Our goal was to ensure practical applicability in machine learning and mental health analytics.
- Our research aimed to explore real-world applications and implications of the methodology, providing valuable insights for both academic research and practical implementations in mental health contexts.

1.5 Organization of the Thesis

The thesis sees a systematic approach to the study of topic under research question. The proposed methodology is based upon a comprehensive literature review presented in Chapter 2. In Chapter 3 we present the proposed methodology, focusing on Ensemble Learning, Self-Training and Explainable AI. Chapter 4 discusses the experimentation setting and findings, emphasizing subtle stress pattern capture as well as performance resilience. Chapter 5 addressed contributions and future directions.

Chapter 2

Literature Review

This chapter addresses machine learning applications of mental health monitoring via social media analysis from the literature review. Includes a range of studies on different ML algorithms such as ensemble learning, semi-supervised learning and transparency issues in AI models. These observations reveal the changing environment and threats in a sector; innovative methods, as well as clear models are necessary for using social media data respectively to be able to determine positive indicators of mental health.

Nijhawan et al. [1] employed machine learning (ML), BERT, and Latent Dirichlet Allocation (LDA) techniques to achieve high accuracy in stress detection from social media interactions. Their study suggested exploring the integration of additional health parameters and enhancing the interpretability of LDA results to further advance stress detection methodologies. By focusing on the nuanced interactions on social media platforms, this work provides valuable insights into the complex dynamics of stress expression and detection in online communities. The findings from this study contribute to the development of more effective strategies for monitoring mental health indicators in digital environments.

KVTKN, Prashanth, and Tene Ramakrishnudu [2] proposed a novel semi-supervised stress detection method named 'SMTSD,' which leverages logistic regression and sarcasm information. Their approach outperformed existing techniques, addressing the challenge of limited labeled data and underutilization of sarcasm in stress detection. By filling a critical research gap, this study opens up avenues for further exploration in semi-supervised learning approaches for stress detection, suggesting future research directions such as multi-task learning to enhance stress and sarcasm detection concurrently. This work not only enhances our understanding of stress detection in social media discourse but also provides practical implications for improving mental health monitoring strategies.

Nibraas Khan and Nilanjan Sarkar [3] utilized Semi-Supervised Generative Adversarial Networks (GANs) for cost-efficient stress detection using partially labeled physiological data. Their study addressed the challenge of expensive labeling by demonstrating the effectiveness of semi-supervised learning algorithms on the WESAD dataset. However, the research highlighted the need for further investigation into the expensive labeling of physiological data for affective state detection. By exploring cost-efficient methods for stress detection, future studies can contribute to the development of scalable and accessible mental health monitoring solutions.

Han Yu and Akane Sano [4] leveraged unlabeled wearable sensor data for stress detection using data augmentation, auto-encoder pre-training, and semi-supervised learning techniques. While their study demonstrated significant performance improvements on three datasets, it also emphasized the importance of model interpretability. Future work suggested by the research includes investigating personalization and interpretation techniques to enhance stress detection accuracy and usability. By addressing these research gaps, future studies can develop more effective and user-friendly stress monitoring systems for diverse populations.

Syed Tanzeel Rabani, Qamar Rayees Khan, and Akib Mohi Ud Din Khanday [5] analyzed suicidal ideation on Twitter using manual annotation and machine learning techniques such as Random Forest. While their study acknowledged the power of prediction uncertainty, it also identified several areas for future research, including incorporating questionnaire inputs, analyzing user connectivity, exploring blog posts, and investigating deep learning algorithms for multi-class classification. By addressing these research gaps, future studies can advance our understanding of suicidal behavior detection on social media platforms and improve intervention strategies to support at-risk individuals.

In their study, Luna Ansari et al. [6] trained text classifiers for depression detection, demonstrating that ensemble models outperformed individual classifiers. Despite their findings, the research identified a need for further exploration, including experimenting with different text classification models and incorporating features like Part-Of-Speech (POS) tags for enhanced performance. By addressing this research gap, future studies can contribute to the development of more robust and accurate depression detection systems, ultimately improving mental health monitoring and intervention strategies.

F. D. Martino and F. Delmastro [7] utilized ensemble learners and Recurrent Neural Networks (RNNs) on the WESAD dataset for high-resolution stress detection. While their work demonstrated advancements in physiological stress prediction, it identified a research gap in exploring the impact of COVID-19 on older adults' stress levels and the effectiveness of transfer learning in stress detection tasks. This study highlights the need for further investigation into the intersection of stress monitoring and global health crises, as well as the potential benefits of leveraging transfer learning techniques for improved stress detection models.

Webb, Geoffrey I., and Zijian Zheng [8] explored multistrategy ensemble learning to improve accuracy by increasing diversity. Their study developed techniques with higher accuracy and highlighted the need for further investigation with different base learners and ensemble sizes. By addressing the trade-off between diversity and test error, future studies can optimize ensemble learning strategies for better performance across diverse datasets and applications.

Uban et al. [9] explored the explainability of depression detection on social media platforms using neural network models and linguistic features. Their study identified a gap in the limited exploration of explainability in mental disorder detection and highlighted the potential benefits of multimodal approaches. By investigating the explainability of depression detection models, future studies can enhance our understanding of model behavior and improve the interpretability of mental health monitoring systems in digital environments.

Angelov, Plamen P., et al. [10] conducted an analytical review of explainable artificial intelligence (XAI), emphasizing the importance of XAI and outlining its applications, challenges, and future directions. The study identified challenges in XAI and stressed the importance of user rights and trustworthy AI systems. By providing a comprehensive analysis of XAI, this work serves as a valuable resource for researchers and practitioners interested in developing transparent and interpretable AI systems.

Arrieta, Alejandro Barredo, et al. [11] conducted a literature review on explainable artificial intelligence (XAI), providing a taxonomy of recent contributions and discussing challenges and opportunities for responsible AI. The study highlighted the need for a comprehensive understanding of the potential and limitations of XAI techniques. By addressing the challenges and opportunities in XAI, this work lays the foundation for developing responsible and trustworthy AI systems.

Quader et al. [12] conducted a comprehensive study spanning from 2017 to 2021, focusing on utilizing various machine learning, ensemble, and deep learning techniques to classify depressive posts on Reddit. Their research showcased the effectiveness of models like BERT and LSTM in accurately identifying depressive content, especially in recent data. The study addressed a notable research gap, highlighting the scarcity of recent, advanced studies on depression detection, with traditional methods dominating the field. By emphasizing the need for advanced models such as BERT and LSTM for more context-aware detection, this work sets a foundation for future research in leveraging sophisticated techniques for mental health monitoring on social media platforms.

Osman et al. [13] conducted a comprehensive review of 52 studies employing machine learning (ML) and deep learning (DL) techniques in mental health analysis, with a specific focus on Support Vector Machines (SVM) and Convolutional Neural Networks (CNN). Despite significant progress in the field, the study identified persistent challenges, including the incorporation of diverse criteria and biases arising from the reliance on social media data. By critically evaluating existing methodologies, this work provides valuable insights into the current landscape of ML and DL applications in mental health analysis, guiding future research towards addressing key challenges and improving the efficacy of computational approaches for mental disorder detection and monitoring.

Alban Maxhuni et al. [14] employed semi-supervised learning, ensemble methods, and transfer learning for stress modeling using scarce smartphone data. While their work demonstrated promising results, it also identified a research gap in understanding the detrimental effects of transfer learning from dissimilar subjects. This study emphasizes the need for further exploration to enhance our understanding of transfer learning techniques and their implications for stress prediction models, ultimately improving the accuracy and reliability of stress monitoring systems in real-world settings.

Huijie Lin et al. [15] developed a hybrid model (FGM + CNN) for stress detection on social media platforms. Their study found a higher number of sparse social structures among stressed users, highlighting the importance of understanding social dynamics in stress detection. However, the research identified a gap in understanding the reasons for social structure differences and their implications for stress detection. By addressing this research gap, future studies can improve the accuracy and reliability of stress detection models by incorporating social network analysis techniques.

Xin Wang et al. [16] proposed a meta-learning-based stress category detection framework using a Dependency Graph Convolutional Network, Mixture of Experts, and Relation Module. Their study addressed the challenge of detecting rarely appeared stress categories on social media platforms without relying on lexicon-based methods. By demonstrating strong generalization capabilities with limited labeled data, this research provides valuable insights into the development of robust stress category detection models. Future studies can build upon these findings to improve the accuracy and scalability of stress monitoring systems in digital environments.

Banerjee et al. [17] conducted an analysis of stress among tech professionals using explainable AI (XAI) techniques. While their study provided valuable insights into the factors contributing to work-related stress, it also identified a research gap in the lack of reported research on flexible systems using XAI to identify work stressors affecting tech professionals. By addressing this gap, future studies can develop more effective and user-centered AI systems for identifying and mitigating work-related stress in the tech industry.

Dong, Xibin, et al. [18] conducted a survey on ensemble learning approaches, acknowledging the need for more efforts to enhance ensemble model performance, especially with complex pattern data. The study identified several challenges in ensemble learning and suggested further research to address these challenges. By providing a comprehensive overview of ensemble learning methodologies, this work serves as a valuable resource for researchers and practitioners interested in leveraging ensemble techniques for various applications.

Zhou, Zhi-Hua [19] presented an overview of various ensemble learning techniques, focusing on presenting ensemble learning methodologies. The study provided insights into the principles and applications of ensemble learning, serving as a valuable reference for researchers and practitioners interested in leveraging ensemble techniques for machine learning tasks.

Seeger, Matthias [20] proposed a taxonomy for categorizing semi-supervised learning methods, addressing the need for a systematic classification. By providing a structured framework for understanding semi-supervised learning techniques, this work facilitates the comparison and evaluation of different approaches, guiding researchers in selecting the most appropriate methods for their specific applications.

Hady, Mohamed Farouk Abdel, Schwenker [21] discussed various aspects of semi-supervised learning, providing insights into the principles, algorithms, and applications of semi-supervised learning. While the study did not specify a particular research gap, it contributed to the understanding of semi-supervised learning techniques and their potential implications for machine learning tasks.

Zou, Yang, et al. [22] proposed a Confidence Regularized Self-Training (CRST) framework for unsupervised domain adaptation, addressing the research gap in exploring novel regularizations for improved inductive biases in unsupervised domain adaptation and semi-supervised learning problems. By introducing a novel regularization technique, this work extends the capabilities of self-training methods and provides insights into enhancing the generalization performance of machine learning models in domain adaptation scenarios.

Vilone, Giulia, and Luca Longo [23] conducted a systematic review of XAI studies, highlighting the lack of consensus on the definition and assessment of explanation validity. The study emphasized the importance of developing standardized evaluation approaches for XAI methods. By addressing the challenges in XAI evaluation, this work contributes to the advancement of transparent and interpretable AI systems.

Dwivedi, Rudresh, et al. [24] conducted a survey on state-of-the-art programming techniques for explainable AI (XAI), emphasizing the importance of transparency and interpretability in intelligent machines. The study discussed various XAI techniques and their applications across different domains. By providing insights into core ideas and techniques in XAI, this work facilitates the development of transparent and interpretable AI systems for real-world applications.

Došilović, Filip Karlo, Mario Brčić, and Nikica Hlupić [25] reviewed recent developments in explainable artificial intelligence (XAI), addressing trust-related problems and incomplete formalization of complex notions. The study highlighted the importance of transparency and interpretability in AI systems and proposed directions for future research. By providing a comprehensive overview of XAI, this work contributes to the advancement of transparent and trustworthy AI systems.

Nadhila Nurdin et al. [26] proposed Bidirectional LSTM and Bi-GRU-LSTM-CNN models for Sentiment Analysis, analyzing using XAI methods. The research suggests exploring pre-trained word embeddings for improved sentiment extension. This study lays the foundation for enhancing the interpretability of NLP models and understanding the underlying factors influencing model predictions.

NADIA FELIX F. DA SILVA, LUIZ F. S. COLETTA, and EDUARDO R. HRUSCHKA [27] conducted a comprehensive survey and comparative study of Semi-Supervised Learning (SSL) approaches for tweet sentiment analysis. The research identified challenges and proposed future research directions. This work provides valuable insights into the application of semi-supervised learning techniques for sentiment analysis tasks, guiding future research in this domain.

Xin Wang et al. [28] developed a stress-specific word embedding learning framework based on the pre-trained language model BERT. The study addressed the need for better stress-specific embeddings by leveraging BERT and proposed self-supervised contrastive learning tasks for enhanced sensitivity to stressors and stressful emotions. By improving the representation of stress-related concepts in word embeddings, this research enhances the performance of stress detection models, particularly in scenarios with limited labeled data.

In summary, these papers collectively offer a comprehensive view of the current state and advancements in machine learning, ensemble learning, and XAI in the context of mental health monitoring. They underscore the importance of innovative methodologies to address data limitations, the need for transparent and explainable models, and the potential of social media as a rich source of mental health indicators.

Chapter 3

Methodology

3.1 Introduction

In this study, a structured approach is used to analyze a dataset. Using the DataFrame, we performed data preprocessing, conduct Word2Vec for text analysis and then refine these models by applying Self-Training as well as a Voting Classifier. Explainable artificial intelligence increases interpretability of models leading to subtle conclusions.

3.2 Experimental Setup

Integrated Development Environment (IDE): We implemented our suggested concept using Google Pro, which is the premium version of Google Colab. It gives high runtime on the one hand and common runtimes for shaping. The experiment was then run on a PC with Windows 11 installed, an 11th Generation Intel Core i5-1135G7 processor clocked at 2.40 and 2.42 GHz, 8.00 GB of RAM, and a 64-bit operating system. Our programming language was Python.

3.3 Data Preprocessing

After reading the CSV file, there are some data preparation operations performed on it so that we can have clean and transformed raw data. To ensure that the work is easy and efficient, we pre-process data after extraction using different NLP tools. But some data don't play any role in stress detection since they only imply a process that is more repetitive. In such a case, preprocessing becomes an important concern.

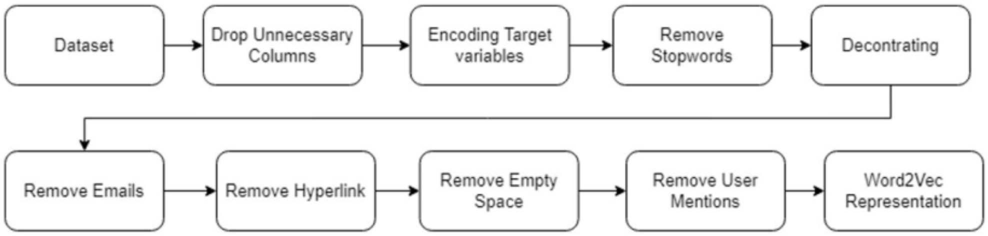

Fig. 3.1. Data preprocessing steps

Drop Unnecessary Columns: The first datasets were checked, and columns that seemed useless or irrelevant for the stress detection analysis process were then dropped to clear up data in order to simplify it.

Encoding Target Variables: The target variables were converted into the numerical values that symbolize stress levels. This encoding is a step that plays an important role in helping machine learning algorithms understand the stress labels.

Remove Stopwords: The natural language process removed common stopwords such as "the", "and", "an", and "is", were irrelevant to meaningful content. This step serves as a noise-reducing measure to improve stress detection models' accuracy.

Decontraction: The text was expanded to include contractions like "won't" to "will not" and "can't" to "cannot". The process of decontraction ensures consistency in word representation and standardizes the language.

Remove Emails: Emails that appeared in the patient information were deleted. Not being associated with stress pattern identification, email addresses as variables are extraneous and that is why their omission simplifies the dataset.

Remove Hyperlinks: Text hyperlink was not included. These links frequently have no relation to stress detection and may contain noise that affects the accuracy of models.

Remove Empty Space: Any remaining gaps or unnecessary random spaces in the text were deleted. This step monitors the consistency of data and improves stress model recognition patterns.

Remove User Mentions: The User mentions, like "@username", were removed instances. Mentions of users could be regarded as noise in the dataset, and they might not make much contribution to stress detection.

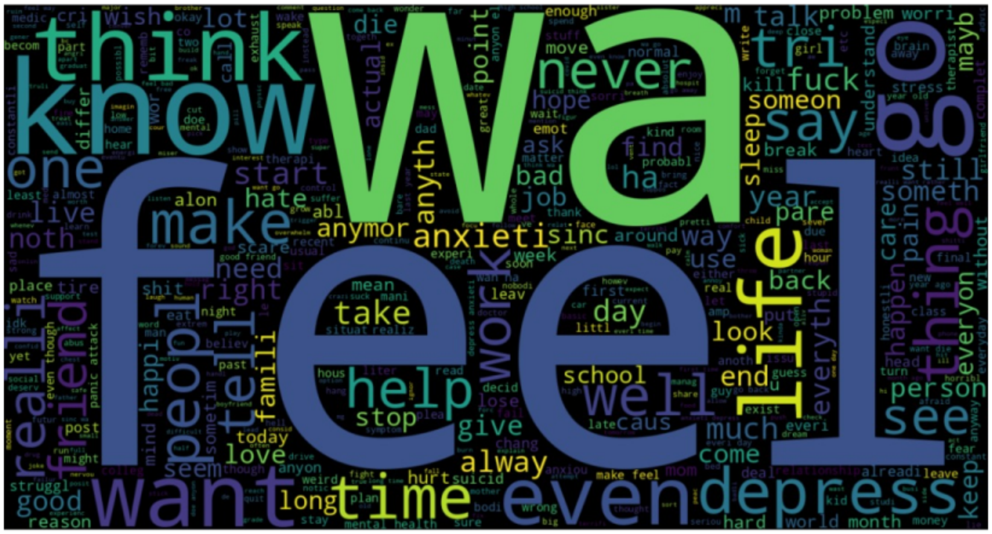

Fig. 3.2. Stress Word Visualization of a dataset

3.4 Feature Extraction

The transformation of raw textual data into meaningful features is crucial for effective stress detection. Two key techniques were employed for feature extraction: Bigrams and Word Embeddings (Word2Vec).

Bigrams: Bigrams capture the co-occurrence of consecutive words in the text, represented as $B_i(w_i, w_{i+1})$. The bigram set B is defined as

$$B = \{B_i(w_i, w_{i+1})\} \tag{3.1}$$

Incorporating bigrams enhances the model's sensitivity to specific word combinations or phrases associated with stress, providing a localized understanding of language patterns.

Word Embeddings (Word2Vec): Word2Vec represents words as dense vectors in a continuous vector space. The embedding function W for a word w is defined as

$$W(w) \rightarrow R^d \tag{3.2}$$

where d is the dimensionality of the vector space. The Word2Vec representations enable the model to capture semantic relationships and contextual meanings of words.

3.5 Feature Selection

Feature selection involved filtering out irrelevant or redundant features to optimize model performance. Let X represent the set of features extracted from both bigrams and Word2Vec representations. The selected feature set X' is obtained through a feature selection function FS

$$X' = FS(X) \tag{3.3}$$

The goal of feature selection is to retain the most informative features while discarding less relevant ones.
By combining localized insights from bigrams and contextual understanding from Word2Vec, our feature extraction and selection process ensures a comprehensive representation of textual content. This strategic approach enhances the subsequent stages of the stress detection model, contributing to its accuracy and effectiveness in deciphering stress signals within social media interactions.

3.6 Model Selection and Training

This section explores the methodology involved in constructing an efficient stress detection model. This involves carefully selecting ensemble models, including Random Forest, Gradient Boosting, Support Vector Machine, k-Nearest Neighbors, Naive Bayes, and Decision Tree. The central aspect of the process revolves around incorporating these classifiers into the innovative EST model. The subsequent stages navigate the complexities of self-training, integrating pseudo-labeling to augment the labeled dataset with confident predictions from the ensemble. This section provides a brief but analytical review of the selected models and the training method, explaining how ensemble learning combined with self-training approaches is brought together.

3.6.1 Voting Ensemble Classifier:

The Voting Ensemble Classifier represents a sophisticated fusion of multiple classification models, each offering its unique perspective on the data. This ensemble approach capitalizes on the diversity among the base classifiers to enhance the overall predictive performance. The ensemble's strength lies in its ability to mitigate the weaknesses of individual classifiers while leveraging their collective intelligence.

In building a Voting Ensemble, the selection of base classifiers is paramount. Each classifier should exhibit complementary strengths and weaknesses, ensuring that they capture diverse aspects of the data distribution. For example, a combination of decision tree-based algorithms like RandomForest and GradientBoosting with linear classifiers like Support Vector Machines (SVM) or logistic regression can offer a rich blend of model characteristics, leading to improved generalization and robustness.

Furthermore, the aggregation strategy employed in the voting process can significantly impact the ensemble's performance. While simple majority voting is often effective, more sophisticated approaches, such as weighted voting or stacking, can further optimize the decision-making process by assigning different weights to each base classifier based on their individual performance metrics or domain expertise.

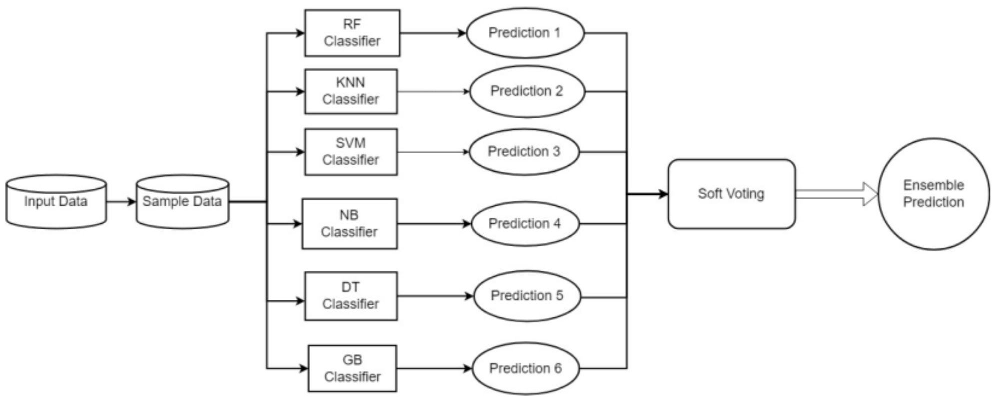

Fig. 3.3. Soft Voting Ensemble Classifier

To maintain diversity and prevent overfitting, techniques like bagging and boosting can be applied. Bagging involves training multiple instances of the same base classifier on different subsets of the training data and aggregating their predictions, thereby reducing variance and improving stability. Boosting, on the other hand, focuses on sequentially training weak learners, with each subsequent learner aiming to correct the errors of its predecessors, resulting in a more accurate final model.

The versatility of the Voting Ensemble Classifier makes it well-suited for various machine learning tasks, ranging from classification and regression to anomaly detection and ensemble pruning. Its ability to harness the collective wisdom of diverse classifiers not only enhances predictive performance but also fosters interpretability and trust, crucial aspects in real-world applications where model decisions impact human lives and decision-making processes.

3.6.2 Self-Training:

Self-Training, also known as self-supervised learning or pseudo-labeling, represents a paradigm shift in supervised learning by embracing the vast amounts of unlabeled data available in many real-world scenarios. The core principle behind self-training is to iteratively refine a model's predictions on unlabeled data and incorporate them into the training process to improve its performance.

At the heart of the self-training loop lies the generation of pseudo-labels for unlabeled data. These pseudo-labels are essentially model predictions, treated as ground truth labels during the training phase. However, the reliability of pseudo-labels can vary significantly, posing a significant challenge in the self-training framework.

To address this challenge, various strategies can be employed to filter out unreliable pseudo-labels. Confidence thresholds, for instance, can be used to accept only those pseudo-labels for which the model exhibits a high level of confidence. Alternatively, ensemble-based approaches, where pseudo-labels are validated across multiple models or iterations, can help improve the robustness of the self-training process.

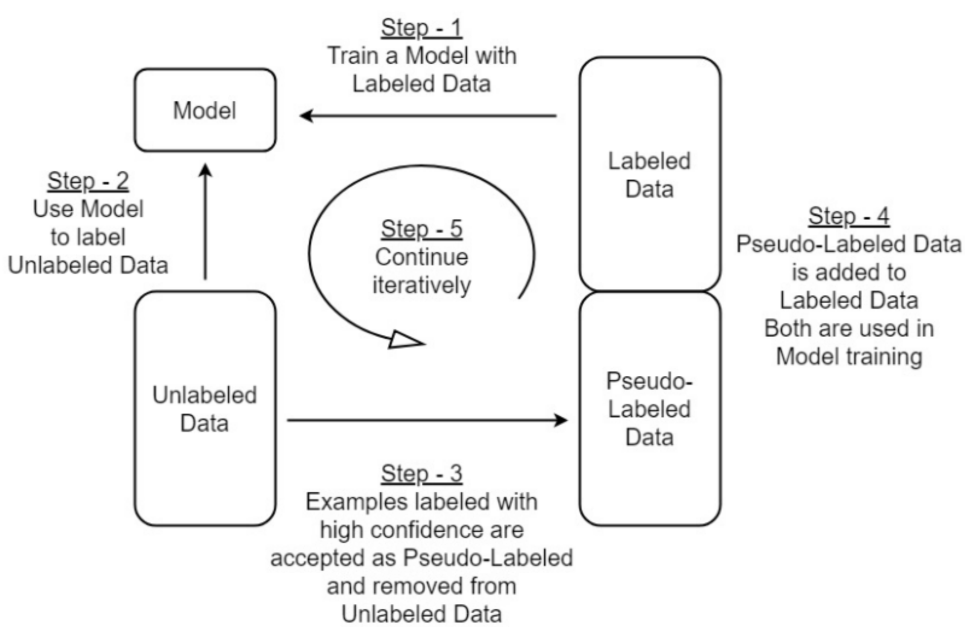

Fig. 3.4. Self Training

Another critical aspect of self-training is the augmentation of the training dataset with pseudo-labeled instances. By combining labeled and pseudo-labeled data, the model gains access to a more diverse and representative sample of the underlying data distribution, leading to improved generalization and robustness.

The iterative nature of self-training allows the model to continuously refine its predictions and adapt to evolving data distributions over time. Through multiple training iterations, the model gradually learns to generalize better, resulting in improved performance on the target task.

Despite its effectiveness, self-training is not without limitations. The reliance on pseudo-labels introduces the risk of error propagation, where incorrect labels may propagate through subsequent iterations, leading to model degradation. Additionally, self-training requires careful tuning of hyperparameters and validation strategies to ensure optimal performance and prevent overfitting.

3.6.3 Explainable AI (XAI):

Explainable AI (XAI) encompasses a diverse set of techniques and methodologies aimed at enhancing the interpretability, transparency, and accountability of machine learning models. In an era where AI systems are increasingly integrated into critical decision-making processes, XAI plays a crucial role in building trust and understanding among users and stakeholders.

Feature importance analysis is a foundational component of XAI, providing insights into the relative contribution of input features to the model's predictions. By quantifying feature importance scores, stakeholders can identify influential features, discern relevant patterns in the data, and validate the model's decision-making process. Here is an example:

In addition to feature importance analysis, visualization techniques such as partial dependence plots and individual conditional expectation (ICE) plots offer intuitive visual representations of the relationship between input features and model predictions. These visualizations enable stakeholders to explore the effects of individual features on the model's output, uncover complex interactions, and gain deeper insights into the underlying data dynamics.

Local interpretability techniques, such as SHAP (SHapley Additive exPlanations) values, provide fine-grained explanations for individual predictions. By attributing the contribution of each feature to the model's output for a specific instance, SHAP values offer localized insights into the rationale behind the model's decisions, facilitating trust, and understanding among users.

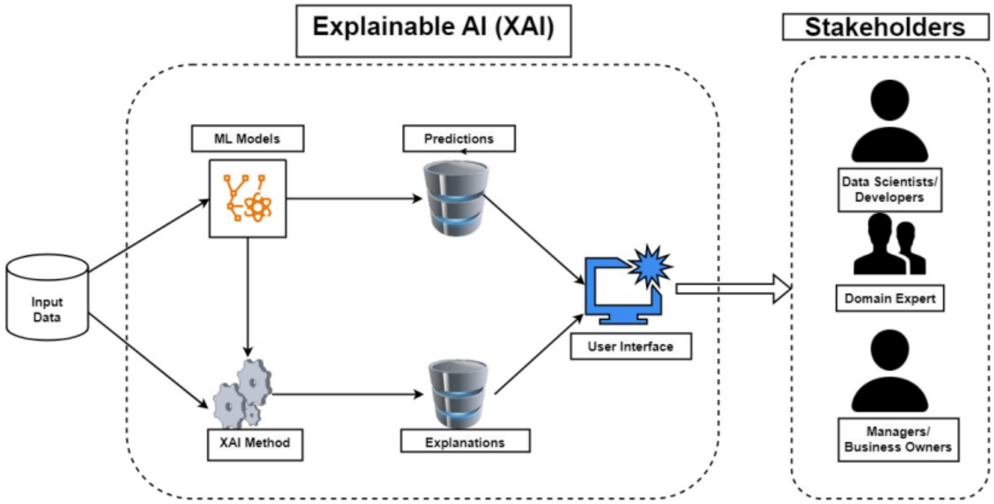

Fig. 3.5. Example of Explainable AI (XAI)

Furthermore, model-agnostic approaches to XAI, such as LIME (Local Interpretable Model-agnostic Explanations), provide interpretable explanations for complex black-box models. By generating locally faithful approximations of model behavior, LIME enables users to understand and validate model predictions across a wide range of applications and domains.

Overall, XAI serves as a bridge between complex machine learning models and end-users, facilitating transparent, accountable, and trustworthy AI systems. By providing interpretable explanations for model predictions, XAI empowers users to make informed decisions, diagnose model behavior, and address potential biases or errors, ultimately enhancing the adoption and acceptance of AI technologies in real-world applications.

3.7 Work Procedure

The research used a systematic methodology of analysis for the data set under investigation. A Data Frame structure was used to load the dataset that involved categorical variables in a specific way using appropriate actions aimed at improving data quality. A number of different text preprocessing approaches were performed that included methods for cleaning up the texts, and filtering mechanisms to identify relevant words in preparation for further analysis.
The use of this word embedding model with the help of Word2Vec was developed and implemented. This was achieved by training the Word2Vec model on the dataset and via Vectors produced from a trained Word2Vec Model, which deepens understanding of textual content.
For ease of analysis in the subsequent step, the dataset was divided into labeled and unlabeled data sets. In turn, the Self-Training methodology was implemented on both labeled and unlabeled data to allow models iteratively improve their accuracy.

$$L_t = L \cup U_t \tag{3.4}$$

here L_t is the labeled set at iteration t, L is the initial labeled set, and U_t is the set of pseudolabeled instances selected based on confidence at iteration t.
A complex Voting Classifier was developed based on the partitioned data, with several algorithms used to improve the forecasting power of ensemble model.

$$\hat{y}_{soft} = \frac{\sum_{j=1}^{N} W_j . \hat{y}}{\sum_{j=1}^{N} W_j} \tag{3.5}$$

here N is the number of models. $\hat{y}$ is the prediction of model j, and W_j is the weight assigned to j.
Moreover, Explainable AI methods were used to illuminate the decision-making process of the models and reveal some factors that made predictions as they did.
In general, the methodology employed a comprehensive approach in which text preprocessing was used along with advanced modelling techniques and Explainable AI to extract valuable insights from the dataset for improving model interpretability.

3.8 Schematic Diagram of the Proposed Model:

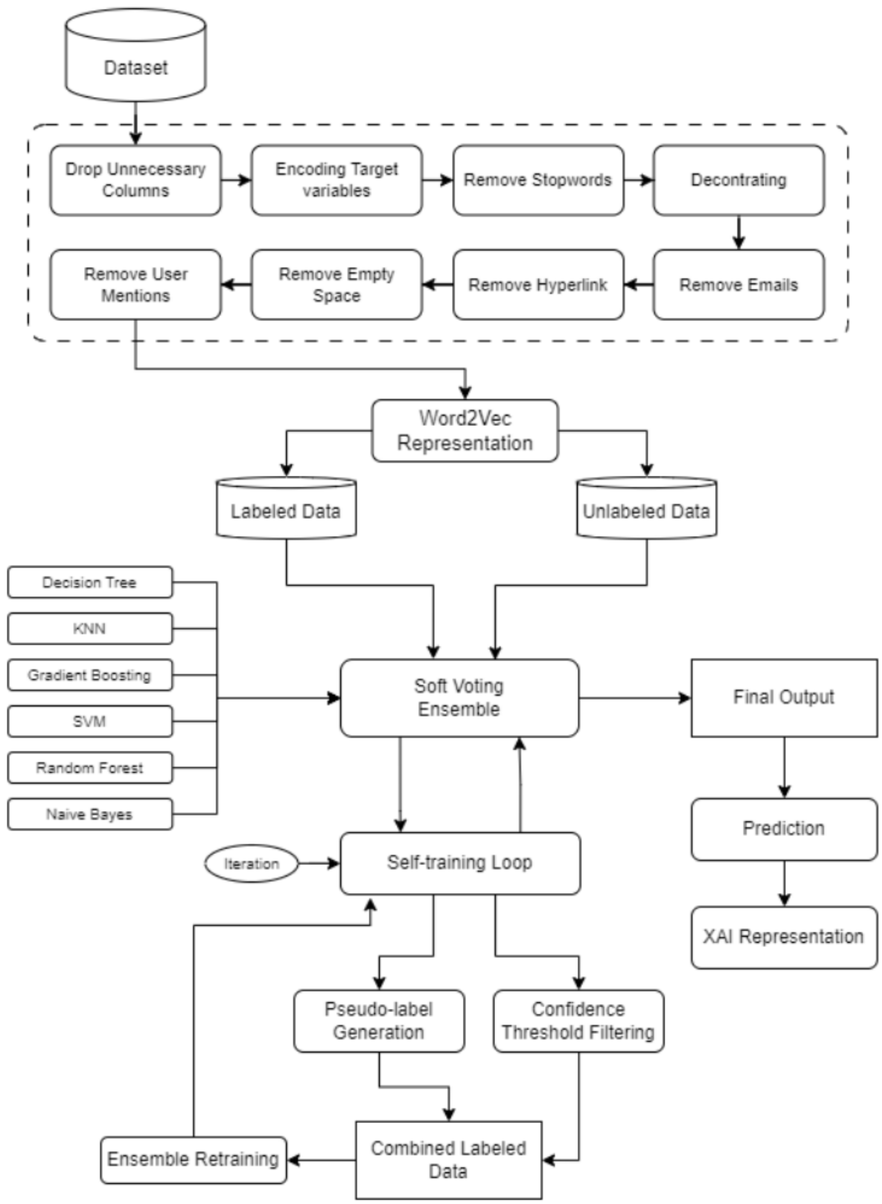

Fig. 3.6. Procedure of EnsembleSelfTrain

3.9 Proposed Algorithm

The algorithm presented in our work follows a sequential approach and consists of the following steps:

Algorithm 1: EnsembleSelfTrain

1. **Require:**
   Whole dataset D; training set ratio α; Ensemble of algorithms (h, h1, h2, h3);
   Iteration times i; Set of labeled training instances L;
   Set of unlabeled training instances U;Confidence threshold C;

2. **Pre-process:**
   Split D into labeled L and unlabeled U;
   U Initialize individual models within an ensemble E;
   Train the ensemble E on the labeled training data L;

3. **Procedure:**

```
for each iteration in the range 1 to i do
    Use the ensemble E to predict pseudo-labels for the unlabeled data U
    Filter pseudo-labels based on the confidence threshold C.
    Combine pseudo-labels with the original labeled data L
    Retrain the ensemble E on the combined data.
end
```

Chapter 4

Results and Discussions

4.1 Introduction

In this chapter, we provide the findings of our research and discuss several issues surrounding them. The section below will elaborate on the predominant conclusions and offer some implications and significance of results. The aim of the discussion is to provide a synthesis between the study outcomes and their broader implications from what was discussed in research objectives, hence offering an overall comprehension on how such contributions can be implemented.

4.2 Dataset Description

The data used in this study comes from Kaggle and concerns conversations on mental health topics via Twitter or Reddit. The labeled and unlabeled Mental Health Corpus from Twitter is large reaching nearly 28,000 tweets. The Human Stress Prediction dataset for Reddit wherein only stress-related language was focused and labeled posts were used for future learning. Dreaddit, another Reddit dataset that strives to be smaller though labeled and able for targeted analysis. The Depression: The Reddit Dataset consists of more than 7,700 posts focusing on stress content with a bigger volume suitable for comprehensive analysis.

Dataset	Data Source	Feature Type	Number of posts	Labeled posts	Unlabeled posts
Mental Health Corpus	Twitter	Textual	27977	5596	22381
Human Stress Prediction	Reddit	Textual	2838	568	2270
Dreaddit	Reddit	Textual	715	143	572
Depression: Reddit Dataset	Reddit	Textual	7731	1546	6185

TABLE 4.1: The details of posts

4.3 Evaluation Metrics:

The Confusion Matrix serves as a pivotal tool in assessing the effectiveness of various machine learning models, aiding in the selection of the most suitable model among those employed. When evaluating a model based not only on accuracy but also on metrics like precision, recall, and F1-score, the Confusion Matrix plays a crucial role. It comprises four key values: True Positive, True Negative, False Positive, and False Negative. The interpretation of these values is elucidated below to calculate Accuracy, Precision, Recall, and F1-score.

True Positive (TP): Indicates how accurately the model classified positive samples as positive. Conversely, False Positive (FP) represents instances where the model incorrectly classified negative samples as positive.

True Negative (TN): Signifies the instances in which the model correctly identified negative samples as negative. False Negatives (FN) denote the frequency with which the model erroneously identified positive samples as negative.

Accuracy: A measure of the percentage of correct predictions, accounting for both positive and negative cases. The formula is given by:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.1}$$

Precision: The ratio of correctly predicted positive cases to all predicted positive cases, demonstrating the accuracy of positive predictions. The formula is expressed as:

$$Precision = \frac{TP}{TP + FP} \tag{4.2}$$

Recall (Sensitivity): Indicates the percentage of positive class instances out of all positive cases identified by the model. It is calculated as:

$$Recall = \frac{TP}{TP + FN} \tag{4.3}$$

F1 Score: A composite metric assessing the model's accuracy based on precision and recall. It is computed by averaging the precision and recall values and is given by:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4.4}$$

The F1 Score becomes particularly relevant when False Positive and False Negative have differing costs, making it a valuable metric to consider alongside Accuracy. The inclusion of both false positives and false negatives in the calculation underscores the F1 Score's comprehensive evaluation of model performance.

4.4 Result Anaylsis

The EnsembleSelfTrain (EST) model, leveraging a diverse ensemble encompassing Random Forest, Gradient Boosting, Support Vector Machine, k-Nearest Neighbors, Naive Bayes, and Decision Tree, coupled with a self-training mechanism, exhibits healthy and consistent performance across various datasets. The iterative self-training process involves predicting pseudo-labels for unlabeled data using Soft Voting ensemble, filtering these pseudo-labels based on a predefined confidence threshold, and merging them with the original labeled data to create an augmented training set. The Soft Voting ensemble model is then retrained on this augmented data, iterating through this process within a specified range.

Dataset	Accuracy (%)	Recall	Precision	F1-Score
Mental Health Corpus	89.29	0.8929	0.9124	0.9025
Human Stress Prediction	89.71	0.8971	0.9146	0.9057
Dreaddit	91.67	0.9167	0.9271	0.9218
Depression: Reddit Dataset	89.52	0.8952	0.8944	0.9044

TABLE 4.2: Results of EST in our datasets

The outcomes, as presented in the tabulated results, showcase notable achievements of the EST model. Consistently high accuracy is observed across all datasets, with Dreaddit reaching an impressive 91.67%. F1-Score values, reflecting a balance between precision and recall, range from 0.892 to 0.914, indicating the model's proficiency in handling both false positives and false negatives. Strong recall values (0.8929 to 0.9167) highlight the model's effectiveness in identifying positive instances, a crucial aspect in stress detection. Precision values (0.8944 to 0.9271) underscore the model's capability to minimize false positives. The accuracy, F1-score, recall, and precision metrics are displayed for the EST model across the four datasets are shown in Figure 4.1

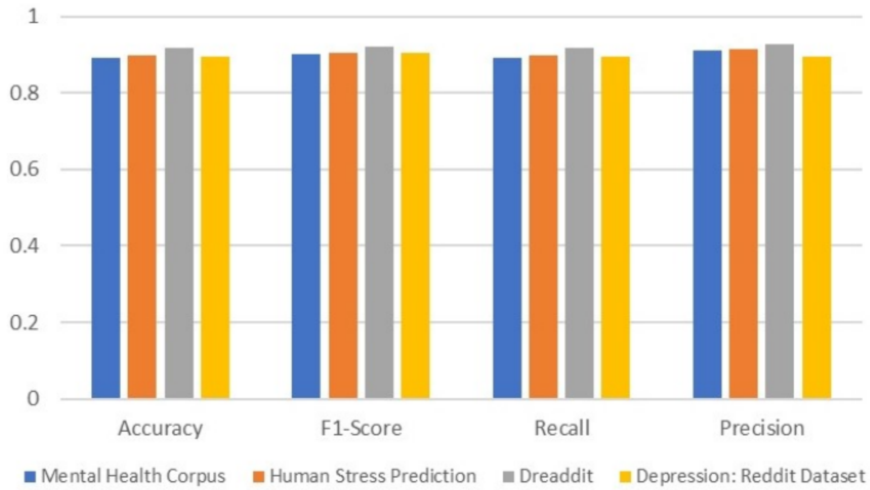

Fig. 4.1. Results of EST in our datasets

The iterative self-training process enhances the model's adaptability, enabling it to iteratively refine predictions. This adaptability is evident in the consistent improvement of performance metrics. With its diverse ensemble and effective self-training strategy, the EST model emerges as a promising approach for stress detection, showcasing versatility and potential for practical applications in the domain of mental health analytics.

Explainable Artificial Intelligence (XAI) Insights:

In our pursuit of ensuring transparency in stress predictions, we implemented Explainable Artificial Intelligence (XAI) techniques to present the underlying logic of the EnsembleSelfTrain (EST) model. This section focuses on providing a glimpse into the interpretability achieved through prediction probabilities and highlighted words.

For each post, the EST model generates prediction probabilities for stress and non-stress categories. These probabilities offer insights into the model's confidence in its predictions. For instance, a probability of 0.77 for "stressed" indicates a high likelihood of the post being categorized as stressed, while 0.23 for "Not stressed" implies a lower confidence in the non-stress classification. This transparency aids users and stakeholders in understanding the model's decision-making process. To further enhance transparency, we present highlighted words that contribute significantly to the model's predictions. In the first example, words like "self-harm", "suicide", "overwhelmed", and "panic attack" carry higher weights, influencing the classification towards stress. Contrastingly, the second example includes words like "encourage," "happiness", and "well-being," which contribute to a non-stress classification.

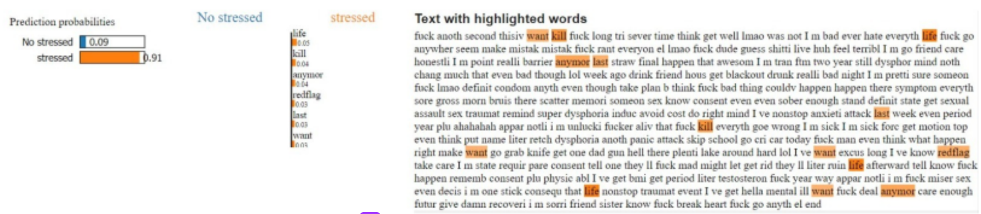

Fig. 4.2. Result of dataset 1's "Stressed" post

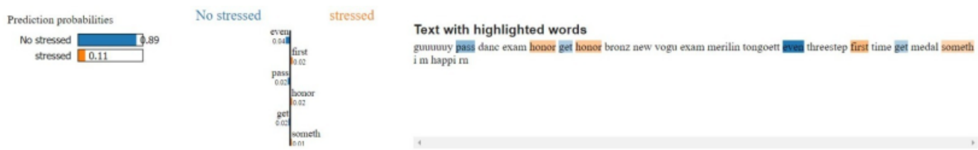

Fig. 4.3. Result of dataset 1's "Not Stressed" post

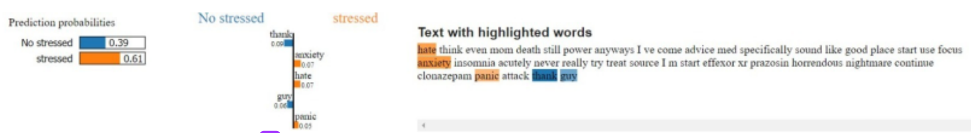

Fig. 4.4. Result of dataset 2's "Stressed" post

Fig. 4.5. Result of dataset 2's "Not Stressed" post

Fig. 4.6. Result of dataset 3's "Stressed" post

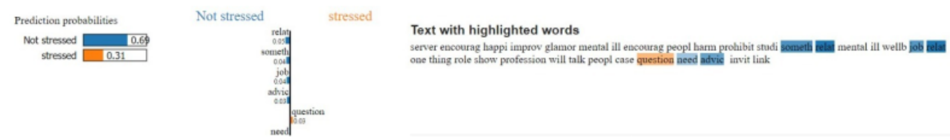

Fig. 4.7. Result of dataset 3's "Not Stressed" post

Fig. 4.8. Result of dataset 4's "Stressed" post

Fig. 4.9. Result of dataset 4's "Not Stressed" post

By combining prediction probabilities and highlighted words, our XAI approach aims to demystify the EST model's decision rationale, empowering users to comprehend and trust the stress predictions made by the system. This transparency is crucial in the context of mental health analytics, fostering user confidence and facilitating meaningful interpretability.

4.5 Comparison with the previous work

In our comprehensive evaluation, the EnsembleSelfTrain (EST) model consistently demonstrated superior performance in stress detection compared to a range of baseline models applied to various datasets. Across distinct datasets such as the Mental Health Corpus, Human Stress Prediction, Dreaddit, and Depression: Reddit Dataset, EST consistently achieved remarkable accuracy, outshining models like Self Training, Logistic Regression, SVM, RF, NB, and Bi-LSTM. Figure 4.10 shows how Baseline models compare to EST in terms of accuracy and F1-score.

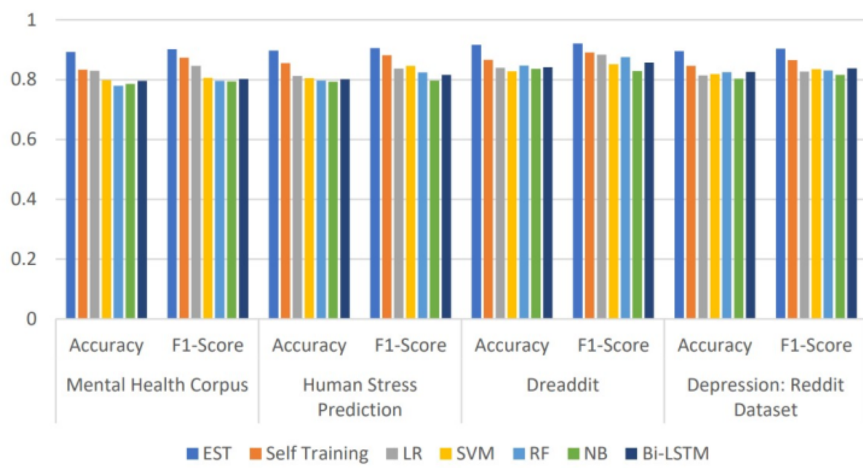

Fig. 4.10. Comparison of Base line models with EST

The results shown in Table 4.3 underscore EST's effectiveness in deciphering stress patterns within social media posts, with accuracy percentages reaching 89.29%, 89.71%, 91.67%, and 89.52% across the respective datasets. Additionally, the F1-Scores of 0.892, 0.896, 0.914, and 0.894 further solidify EST's superiority over alternative models. This consistent trend establishes EST as a strong and reliable tool for stress detection in diverse online contexts.

Model	Mental Health Corpus		Human Stress Prediction		Dreaddit		Depression: Reddit Dataset	
	Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score
EST	**89.29**	**0.9025**	**89.71**	**0.9057**	**91.67**	**0.9218**	**89.52**	**0.9044**
Self Training	83.32	0.874	85.58	0.882	86.61	0.891	84.63	0.865
LR	82.94	0.846	81.25	0.837	83.94	0.884	81.44	0.827
SVM	79.85	0.806	80.54	0.846	82.81	0.852	81.87	0.835
RF	77.97	0.796	79.68	0.824	84.68	0.875	82.49	0.831
NB	78.64	0.794	79.29	0.797	83.66	0.829	80.37	0.816
Bi-LSTM	79.58	0.802	80.14	0.816	84.15	0.857	82.64	0.838

TABLE 4.3: Comparison of Base line models with EST

The comparison with baseline models highlights EST's ability, emphasizing the utility of ensemble model and self-training techniques for enhanced accuracy and reliability in mental health analytics applied to social media discourse. These findings signify a significant advancement in stress detection methodologies, with EST emerging as a promising solution for understanding and addressing mental health concerns within online communities.

Chapter 5

Conclusion

5.1 Research Summary

In conclusion, the research presents a methodology integrating Ensemble Learning, Self-Training, and Explainable Artificial Intelligence (XAI) for stress detection in social media texts. The utilization of Ensemble Learning and Self-Training has shown promise in enhancing accuracy, while the integration of Word Embedding Models with Word2Vec contributes to capturing nuanced stress patterns. The incorporation of XAI techniques further ensures transparency and interpretability, providing valuable insights into the decision-making processes of the stress detection model. Our findings demonstrate the potential of the proposed methodology in advancing stress detection capabilities, marking a significant stride in the intersection of machine learning and mental health analytics. The model's highest accuracy of 91.67% underscores its robust performance in deciphering stress signals within social media interactions.

5.2 Contribution of the work

The following are some contributions of the work:

- In this work, we brought a new methodology to use for stress detection of social media texts based on Ensemble Learning and Self-Training integrated with Explainable Artificial Intelligence (XAI). This novel methodology not only improves accuracy by means of Ensemble Learning and Self-Training but also enables XAI that makes a leap forward in mental health analytics.
- We supplemented by utilizing Word Embedding Models using Word2Vec, allowing the model to underscore subtle stress patterns in social media posts. This addition improved sensitivity to the various agents of stress and increased their information power and ability to interpret signals from within interactions.
- The overall accuracy of our research stood at a remarkable 91.67%, drawing attention to the excellent performance of an integrated approach we followed in this study. This highlighted its ability to support real-world applications, in such a way that it is essentially situated at the intersection between machine learning and mental health analytics.

5.3 Future Work

Future endeavors could focus on expanding the dataset to encompass diverse linguistic expressions and cultural differences, ensuring the model's applicability across a broader spectrum. Further refining of the XAI algorithms should also be investigated to provide more detailed insights into individual variables driving stress predictions.

The integration of real-time monitoring and continuous learning mechanisms could be investigated to enhance the model's adaptability to evolving language trends and emerging stress-related expressions on social media platforms. Collaborations with mental health professionals and user feedback incorporation may also enrich the model's practical utility and ensure its alignment with the evolving landscape of mental health awareness in digital spaces. Continued research in this direction holds the potential for even more effective and socially responsible stress detection solutions.

References

[1] T. Nijhawan, G. Attigeri, and T. Ananthakrishna, "Stress detection using natural language processing and machine learning over social interactions," *Journal of Big Data*, vol. 9, no. 1, pp. 1–24, 2022.

[2] P. KVTKN and T. Ramakrishnudu, "Semi-supervised approach for tweet-level stress detection," *Natural Language Processing Journal*, p. 100019, 2023.

[3] N. Khan and N. Sarkar, "Semi-supervised generative adversarial network for stress detection using partially labeled physiological data," *arXiv preprint arXiv:2206.14976*, 2022.

[4] H. Yu and A. Sano, "Semi-supervised learning and data augmentation in wearable-based momentary stress detection in the wild," *arXiv preprint arXiv:2202.12935*, 2022.

[5] S. T. Rabani, Q. R. Khan, and A. M. U. D. Khanday, "Detection of suicidal ideation on twitter using machine learning & ensemble approaches," *Baghdad science journal*, vol. 17, no. 4, pp. 1328–1328, 2020.

[6] L. Ansari, S. Ji, Q. Chen, and E. Cambria, "Ensemble hybrid learning methods for automated depression detection," *IEEE transactions on computational social systems*, vol. 10, no. 1, pp. 211–219, 2022.

[7] F. Di Martino and F. Delmastro, "High-resolution physiological stress prediction models based on ensemble learning and recurrent neural networks," in *2020 IEEE symposium on computers and communications (ISCC)*, pp. 1–6, IEEE, 2020.

[8] G. I. Webb and Z. Zheng, "Multistrategy ensemble learning: Reducing error by combining ensemble learning techniques," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 8, pp. 980–991, 2004.

[9] A.-S. Uban, B. Chulvi, and P. Rosso, "Explainability of depression detection on social media: From deep learning models to psychological interpretations and multimodality," in *Early Detection of Mental Health Disorders by Social Media Monitoring: The First Five Years of the eRisk Project*, pp. 289–320, Springer, 2022.

[10] P. P. Angelov, E. A. Soares, R. Jiang, N. I. Arnold, and P. M. Atkinson, "Explainable artificial intelligence: an analytical review," *WIREs Data Mining and Knowledge Discovery*, vol. 11, no. 5, p. e1424, 2021.

[11] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," *Information Fusion*, vol. 58, pp. 82–115, 2020.

[12] A. W. Quader, M. S. Jamil, M. I. Uddin, and M. J. Patwary, "A machine-learning approach to classify depressive posts from reddit social media," in *2023 International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM)*, pp. 1–6, IEEE, 2023.

[13] A. B. Osman, F. Tabassum, M. J. Patwary, A. Imteaj, T. Alam, M. A. S. Bhuiyan, and M. H. Miraz, "Examining mental disorder/psychological chaos through various ml and dl techniques: a critical review," *Annals of Emerging Technologies in Computing (AETiC)*, pp. 61–71, 2022.

[14] A. Maxhuni, P. Hernandez-Leal, L. E. Sucar, V. Osmani, E. F. Morales, and O. Mayora, "Stress modelling and prediction in presence of scarce data," *Journal of biomedical informatics*, vol. 63, pp. 344–356, 2016.

[15] H. Lin, J. Jia, J. Qiu, Y. Zhang, G. Shen, L. Xie, J. Tang, L. Feng, and T.-S. Chua, "Detecting stress based on social interactions in social networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 9, pp. 1820–1833, 2017.

[16] X. Wang, L. Cao, H. Zhang, L. Feng, Y. Ding, and N. Li, "A meta-learning based stress category detection framework on social media," in *Proceedings of the ACM Web Conference 2022*, pp. 2925–2935, 2022.

[17] J. S. Banerjee, A. Chakraborty, M. Mahmud, U. Kar, M. Lahby, and G. Saha, "Explainable artificial intelligence (xai) based analysis of stress among tech workers amidst covid-19 pandemic," in *Advanced AI and Internet of Health Things for Combating Pandemics*, pp. 151–174, Springer, 2023.

[18] X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma, "A survey on ensemble learning," *Frontiers of Computer Science*, vol. 14, pp. 241–258, 2020.

[19] Z.-H. Zhou and Z.-H. Zhou, *Ensemble learning*. Springer, 2021.

[20] M. Seeger, "A taxonomy for semi-supervised learning methods," *Semi-Supervised Learning*, pp. 15–32, 2006.

[21] M. F. A. Hady and F. Schwenker, "Semi-supervised learning," *Handbook on Neural Information Processing*, pp. 215–239, 2013.

[22] Y. Zou, Z. Yu, X. Liu, B. Kumar, and J. Wang, "Confidence regularized self-training," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 5982–5991, 2019.

[23] G. Vilone and L. Longo, "Notions of explainability and evaluation approaches for explainable artificial intelligence," *Information Fusion*, vol. 76, pp. 89–106, 2021.

[24] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan, and R. Ranjan, "Explainable ai (xai): Core ideas, techniques, and solutions," *ACM Comput. Surv.*, vol. 55, jan 2023.

[25] F. K. Došilović, M. Brčić, and N. Hlupić, "Explainable artificial intelligence: A survey," in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 0210–0215, 2018.

[26] N. Nurdin *et al.*, "Explainable artificial intelligence (xai) towards model personality in nlp task," *IPTEK The Journal of Engineering*, vol. 7, no. 1, pp. 11–15, 2021.

[27] N. F. F. D. Silva, L. F. Coletta, and E. R. Hruschka, "A survey and comparative study of tweet sentiment analysis via semi-supervised learning," *ACM Computing Surveys (CSUR)*, vol. 49, no. 1, pp. 1–26, 2016.

[28] X. Wang, H. Zhang, L. Cao, K. Zeng, Q. Li, N. Li, and L. Feng, "Contrastive learning of stress-specific word embedding for social media based stress detection," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 5137–5149, 2023.

EST__A_Semi_supervised_ensemble_model_for_Explainable...
4.pdf

ORIGINALITY REPORT

13%	9%	7%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	1%
2	Submitted to University College London Student Paper	<1%
3	Yun Li, Jie Xu, Xiaobing Zhang, Chen Li, Yingjuan Zhang. "An Incremental Closed Frequent Itemsets Mining Algorithm Based on Shadow Prefix Tree", 2013 10th Web Information System and Application Conference, 2013 Publication	<1%
4	etd.lib.metu.edu.tr Internet Source	<1%
5	Xin Wang, Huijun Zhang, Lei Cao, Kaisheng Zeng, Qi Li, Ningyun Li, Ling Feng. "Contrastive Learning of Stress-specific Word Embedding for Social Media based Stress Detection", Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023	<1%

Publication

6	Ashutosh Anshul, Gumpili Sai Pranav, Mohammad Zia Ur Rehman, Nagendra Kumar. "A Multimodal Framework for Depression Detection During COVID-19 via Harvesting Social Media", IEEE Transactions on Computational Social Systems, 2023 Publication	<1 %
7	www.biorxiv.org Internet Source	<1 %
8	www.coursehero.com Internet Source	<1 %
9	Submitted to Nepal College of Information Technology Student Paper	<1 %
10	umpir.ump.edu.my Internet Source	<1 %
11	uwspace.uwaterloo.ca Internet Source	<1 %
12	www.termpaperwarehouse.com Internet Source	<1 %
13	Sahraoui Dhelim, Liming Chen, Sajal K. Das, Huansheng Ning et al. "Detecting Mental Distresses Using Social Behavior Analysis in the Context of COVID-19: A Survey", ACM Computing Surveys, 2023 Publication	<1 %

14	ebin.pub Internet Source	<1 %
15	onlinelibrary.wiley.com Internet Source	<1 %
16	Nura Shifa Musa, Nada Masood Mirza, Saida Hafsa Rafique, Amira Abdallah, Thangavel Murugan. "Machine Learning and Deep Learning Techniques for Distributed Denial of Service Anomaly Detection in Software Defined Networks - Current Research Solutions", IEEE Access, 2024 Publication	<1 %
17	Submitted to University of New England Student Paper	<1 %
18	d197for5662m48.cloudfront.net Internet Source	<1 %
19	repository.nwu.ac.za Internet Source	<1 %
20	Submitted to University of Strathclyde Student Paper	<1 %
21	export.arxiv.org Internet Source	<1 %
22	mafiadoc.com Internet Source	<1 %

23	Tumlumbe Juliana Chengula, Judith Mwakalonge, Gurcan Comert, Saidi Siuhi. "Improving Road Safety with Ensemble Learning: Detecting Driver Anomalies Using Vehicle Inbuilt Cameras", Machine Learning with Applications, 2023 Publication	<1%
24	archive.org Internet Source	<1%
25	dspace.daffodilvarsity.edu.bd:8080 Internet Source	<1%
26	www.ijritcc.org Internet Source	<1%
27	www.scilit.net Internet Source	<1%
28	Submitted to Cardiff University Student Paper	<1%
29	Submitted to The Robert Gordon University Student Paper	<1%
30	citeseerx.ist.psu.edu Internet Source	<1%
31	www.grin.com Internet Source	<1%
32	Han Yu, Akane Sano. "Semi-Supervised Learning for Wearable-based Momentary	<1%

Stress Detection in the Wild", Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2023
Publication

33	Submitted to University of Lancaster Student Paper	<1 %
34	www.cureus.com Internet Source	<1 %
35	"Research on e-Learning and ICT in Education", Springer Science and Business Media LLC, 2018 Publication	<1 %
36	Submitted to IIT Delhi Student Paper	<1 %
37	Tianlin Zhang, Kailai Yang, Shaoxiong Ji, Sophia Ananiadou. "Emotion fusion for mental illness detection from social media: A survey", Information Fusion, 2022 Publication	<1 %
38	Adib Wahid Quader, Md Shafayet Jamil, Mohammad Irfan Uddin, Muhammed J. A. Patwary. "A Machine-Learning Approach to Classify Depressive Posts from Reddit Social Media", 2023 International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM), 2023 Publication	<1 %

39	Ana-Maria Bucur, Berta Chulvi, Adrian Cosma, Paolo Rosso. "The Expression of Happiness in Social Media of Individuals Diagnosed With Depression", Institute of Electrical and Electronics Engineers (IEEE), 2023 Publication	<1%
40	Yuming Li, Johnny Chan, Gabrielle Peko, David Sundaram. "An explanation framework and method for AI-based text emotion analysis and visualisation", Decision Support Systems, 2024 Publication	<1%
41	dokumen.pub Internet Source	<1%
42	iptek.its.ac.id Internet Source	<1%
43	journals.plos.org Internet Source	<1%
44	oa.upm.es Internet Source	<1%
45	scholar.archive.org Internet Source	<1%
46	www.frontiersin.org Internet Source	<1%
47	Dayan Guan, Jiaxing Huang, Shijian Lu, Aoran Xiao. "Scale Variance Minimization for	<1%

Unsupervised Domain Adaptation in Image Segmentation", Pattern Recognition, 2020
Publication

48 Submitted to University of Essex <1%
Student Paper

49 Weihao Ren. "A novel approach for automatic detection and identification of inappropriate postures and movements of table tennis players", Soft Computing, 2024 <1%
Publication

50 cse.iiuc.ac.bd <1%
Internet Source

51 dspace.univ-adrar.edu.dz <1%
Internet Source

52 easychair.org <1%
Internet Source

53 nozdr.ru <1%
Internet Source

54 papyrus.bib.umontreal.ca <1%
Internet Source

55 peerj.com <1%
Internet Source

56 rinacional.tecnm.mx <1%
Internet Source

57 www.acemap.info
Internet Source

<1 %

58 www.wjgnet.com
Internet Source
<1 %

59 Chaitanya B Pednekar, P. Saravanan. "A Study on Different Methods in Sentiment Analysis from Text", 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA), 2023
Publication
<1 %

60 G.I. Webb, Z. Zheng. "Multistrategy ensemble learning: reducing error by combining ensemble learning techniques", IEEE Transactions on Knowledge and Data Engineering, 2004
Publication
<1 %

61 Giovanni Annuzzi, Andrea Apicella, Pasquale Arpaia, Lutgarda Bozzetto et al. "Exploring Nutritional Influence on Blood Glucose Forecasting for Type 1 Diabetes Using Explainable AI", IEEE Journal of Biomedical and Health Informatics, 2023
Publication
<1 %

62 Lecture Notes in Computer Science, 2006.
Publication
<1 %

63 Submitted to Manchester Metropolitan University
<1 %

Student Paper

64	Rishi Narang, Upendra Pratap Singh. "Interpretable Sequence Models for the Sales Forecasting Task: A Review", 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), 2023 Publication	<1 %
65	academic.oup.com Internet Source	<1 %
66	arxiv.org Internet Source	<1 %
67	doras.dcu.ie Internet Source	<1 %
68	dr.ntu.edu.sg Internet Source	<1 %
69	ir.bdu.edu.et Internet Source	<1 %
70	rucore.libraries.rutgers.edu Internet Source	<1 %
71	s-space.snu.ac.kr Internet Source	<1 %
72	trexvis.github.io Internet Source	<1 %
73	www.arxiv-vanity.com	

Internet Source

<1%

74 www.research-collection.ethz.ch
Internet Source

<1%

75 Amina Adadi, Mohammed Berrada. "Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI)", IEEE Access, 2018
Publication

<1%

76 José Miguel Ramírez-Sanz, Jose-Alberto Maestro-Prieto, Álvar Arnaiz-González, Andrés Bustillo. "Semi-supervised learning for industrial fault detection and diagnosis: A systemic review", ISA Transactions, 2023
Publication

<1%

77 Marko Pregeljc, Erik Štrumbelj, Miran Mihelcic, Igor Kononenko. "chapter 12 Learning and Explaining the Impact of Enterprises' Organizational Quality on their Economic Results", IGI Global, 2012
Publication

<1%

78 Nadia Felix F. Da Silva, Luiz F. S. Coletta, Eduardo R. Hruschka. "A Survey and Comparative Study of Tweet Sentiment Analysis via Semi-Supervised Learning", ACM Computing Surveys, 2016
Publication

<1%

79	Sanchari Saha, R N Sree Vaishnavi, Tejaswi B N. "Application of Factor Graph Model in Stress Detection Using Social Network Messages", IOS Press, 2024 Publication	<1 %
80	"Applied Intelligence and Informatics", Springer Science and Business Media LLC, 2022 Publication	<1 %
81	Lecture Notes in Computer Science, 2010. Publication	<1 %
82	Ton Duc Thang University Publication	<1 %
83	Wenji Wang, Qing Xia, Zhiqiang Hu, Zhennan Yan, Zhuowei Li, Yang Wu, Ning Huang, Yue Gao, Dimitris Metaxas, Shaoting Zhang. "Few-shot Learning by a Cascaded Framework with Shape-constrained Pseudo Label Assessment for Whole Heart Segmentation", IEEE Transactions on Medical Imaging, 2021 Publication	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		