IMFinE: An Integrated BERT-CNN-BiGRU Model for Mental Health Detection in Financial Context on Textual Data

Ashraf Kamal **Padmapriya Mohankumar** **Vishal Kumar Singh**
{ashrafkamal.mca,padmapriya.mohankumar,vishalsingh7x}@gmail.com

Abstract

Nowadays, mental health is a global issue. It is a pervasive phenomenon over online social network platforms. It is observed in varied categories, such as *depression*, *suicide*, and *stress* on the Web. Hence, mental health detection problem is receiving continuous attention among computational linguistics researchers. On the other hand, public emotions and reactions play a significant role in financial domain and the issue of mental health is directly associated. In this paper, we propose a new study to detect mental health in financial context. It starts with two-step data filtration steps to prepare the mental health dataset in financial context. A new model called IMFinE is introduced. It consists of an input layer, followed by two relevant BERT embedding layers, a convolutional neural network, a bidirectional gated recurrent unit, and finally, dense and output layers. The empirical evaluation of the proposed model is performed on Reddit datasets and it shows impressive results in terms of *precision*, *recall*, and *f-score*. It also outperforms relevant state-of-the-art and baseline methods. To the best of our knowledge, this is the first study on mental health detection in financial context.

1 Introduction

The popularity of online social network (OSN) platforms, such as Twitter, Facebook, and Reddit have been growing at an unprecedented rate (Khan et al., 2022a). They have become a real-time and large-scale communication source to find and connect with users across the globe (Kamal and Abulaish, 2019a). These platforms offer users to express their moods, emotions, sentiments, and views. Besides that, users are involved in several activities like exchanging messages with friends, participating in ongoing trends, and establishing connection with celebrities. Hence, an enormous amount of data is generated from these online platforms (Abulaish and Kamal, 2018). It contains rich semantic information which can be used for several applications, such as *sentiment analysis* and *opinion mining*, *predictive modeling*, *Web surveillance*, *recommendation systems*, *information retrieval*, *data summarization*, and *cyber-security* (Khan et al., 2022b). In addition, it is also beneficial for several interdisciplinary studies including *psychology*, *behavioral*, and *cognitive* (Kamal and Abulaish, 2022).

In the last few years, it is seen that users consider OSNs to address mental health problems. Especially, after the occurrence of the *coronavirus disease* (COVID-19) pandemic, which has affected us at physical-, mental-, and psychological-level. The number of cases related to mental health problems, such as *depression*, *stress*, *anxiety*, and *suicide* have surged. As a result, it is a briefly discussed topic world-wide and rapidly increasing across OSNs in the form of *tweets*, *posts*, *comments*, and *blogs*. In the latest survey, the rate of increase of such mental health problems is found larger as compared to the physical health impacts in China (Huang and Zhao, 2020). Also, it is seen that around 80% of people reveal their intention of committing suicide on social media platforms (Sawhney et al., 2021). *Depression* causes frequent *stress* and it occurs due to many reasons, such as stressful life events including bereavement, divorce, prolonged illness, or financial issues. Further, prevailing it for a long span of time develops suicidal tendencies (Ansari et al., 2021).

Recently, people have been taking OSN platforms to communicate and receive advice on mental health-related issues. It has also motivated computational linguistic researchers in the sense that information mining from massive amount of user-generated contents (UGCs) can be used in mental health identification and detection. In this line, a wide range of natural language processing (NLP) and data mining approaches

Proceedings of the 19th International Conference on Natural Language Processing (ICON), pages 139 - 148
December 15-18, 2022 ©2022 Association for Computational Linguistics

and techniques are applied along with handcrafted features for mental health classification on UGCs. Therefore, several traditional machine learning (ML) techniques and advanced deep learning (DL) models are taken into consideration for mental health classification (Ameer et al., 2022).

1.1 Mental Health and Finance

Financial stress is considered an economic determinant of mental health problems. It is directly linked with *depression* and *stress*. It spans world-wide, but mainly in those countries that have large populations and low-income (Alanazi et al., 2022). Existing literature has highlighted positive links between *depression* and several factors which lead to financial stress including debt, financial hardship, financial condition, economic situation, poverty, loan, mortgage, or low-income. Panic related to debt via loan like burdens may lead to sleeping disorder problem such as insomnia (Weissberger et al., 2020). It occurs due to the lack of sleep and affects the mental health of an individual. Similarly, other factors like low-income, mortgage, and poverty have affected an individuals' mental health which further leads to *depression* and *suicide* (Fitch et al., 2007). However, minimal attention is received in existing state-of-the-art (SOTA) in this direction of research as of now, especially using DL-based models. To this end, this study presents IMFinE, a new DL-based model to detect mental health in financial context.

1.2 Our Contributions

The role of context is crucial in text classification problems (Abulaish et al., 2020). In addition to that capturing semantic and syntactic information is also important for classification problems in the textual portion (Kamal and Abulaish, 2019b). Extracting mental health-related semantic and syntactic information via social media content is a challenging, interesting, and notable research problem, especially, when it is mentioned in a financial context. In this direction, this study presents a DL model for detecting mental health in financial context. Figure 1 presents the pipeline of the proposed work. It starts with data collection from Reddit, follow by data pre-processing, two-level data filtration steps to filter candidate mental health-based financial text, and finally a new DL-based model IMFinE is introduced which is responsible for binary classification task. IMFinE consists of an input layer, two parallel relevant BERT embeddings (MentalBERT (Ji et al., 2021) and FinBERT (Araci, 2019)), a convolutional neural network (CNN) (Kim, 2014), a bidirectional gated recurrent unit (BIGRU) (Liu and Guo, 2019), and end up with a dense and output layers.

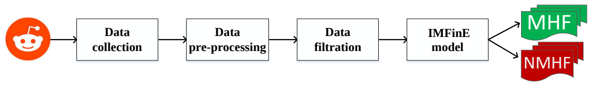


Figure 1: Pipeline of the proposed work

In, IMFinE, the input layer receives input either as a filtered candidate mental health-based financial (MHF) (refer as a positive class) text or normal non-mental health financial (NMFH) (refer as a negative class) text. Thereafter, it is passed to two parallel BERT embeddings (MentalBERT and FinBERT) to retrieve contextual information related to mental health and finance, respectively, and further it is concatenated to give comprehensive context representation. The role of the CNN is to receive that comprehensive context representation and obtain mental health and financial semantic and syntactic information, and the BiGRU layer is used to obtain preceding and succeeding mental health and financial information latent contextual information sequences present in the input text. Thereafter, it is forwarded to the dense and output layer, wherein *sigmoid* activation function is used to classify the input text as either MHF or NMHF.

Overall, the main contributions of this study are as follows:

- Exploring a novel mental health detection on financial textual data.
- Implementation of a two-steps dataset filtration technique to identify candidate mental health-related financial texts.
- Development of a new DL-based IMFinE model to detect mental health in financial context.
- Conducting an empirical evaluation of the proposed model and compared with SOTA and baselines methods to examine its efficacy.

The rest of the paper is organized as follows. Section 2 presents a brief review of the existing works. Section 3 highlights the problem description and provides insights about the dataset preparation.

Section 4 presents the architectural work-flow of the proposed model. Section 5 demonstrates the experimental setup and evaluation results. Finally, Section 6 concludes the paper and highlights future directions of research.

2 Related Work

This section presents the existing SOTA related to mental health identification and detection on OSN platforms. Besides that, this section also highlights the current status and limitations in the end.

Park et al. (2012) analyzed the language which indicates depressive moods on `Twitter` data. They considered depressive attitudes and actions of users, organized and conducted interviews between two users, and analyzed the correlation between interviews and data available on `Twitter`. Choudhury et al. (2013) highlighted "major depressive disorders" prediction and considered many behavioral patterns of depressed users from this social media data. Tsugawa et al. (2015) considered users behavior to predict *depression* on `Twitter` in Japanese language. Ronghua and Qingpeng (2016) explained that users reveal their moods related to depression analysis via social media. Shen et al. (2017) introduced six depression-based feature groups. They proposed a multi-modal *depressive* dictionary learning model for depressed user detection on `Twitter`. Haque et al. (2021) identified *suicide* and *depression* via DL techniques. Xue et al. (2014) proposed stress detection on `Twitter`. They analyzed psychological pressures in teenagers' tweets. Lin et al. (2016) extracted semantic features and combined multi-task learning using CNN for identification of *stress* related topics and events on social media data. Thelwall (2017) introduced a rule-based approach to analyze *stress* and relaxation using both direct and indirect expressions on `Twitter`. Turcan and McKeown (2019) considered `Reddit` and proposed a dataset namely, `Dreadit` for stress classification. They applied a manual approach to construct this dataset. Coppersmith et al. (2016) considered posts by users before the suicide attempt on `Twitter` and analyzed the lexical markers and emotions in it. Giannakakis et al. (2017) considered facial clues from the recorded video for *stress* and *anxiety* detection using ML techniques. Li and Liu (2020) presented stress detection using DL techniques. They applied CNN and multi-layer perceptron. Rastogi et al. (2022) proposed a new dataset for *stress* detection on `Twitter` and `Reddit` datasets.

All aforementioned literature presents the availability of the datasets across OSN platforms and highlights important insights for mental health detection in terms of *depression*, *suicide*, and *stress*. However, there is no literature that addresses mental health detection in financial context using textual data, which is a challenging, worth, and significant research investigation task. To the best of our knowledge, this is the first study on mental health detection in financial context on textual data via DL-based models.

3 Problem Description and Dataset Preparation

This section presents the problem description for mental health detection in finance and covers dataset preparation.

3.1 Problem Description

This study presents the mental health detection in finance on textual data. The considered problem represents a binary classification problem. A piece of textual data is classified as either MHF or NMHF. In this study, we consider three commonly found mental health-related categories – *depression*, *suicide*, and *stress* associated to financial textual data over social media content. A formal definition of each mental health categories is given below followed by a relevant example taken from the `Reddit` posts.

- *Depression*: It is defined as a "*medical condition in which a person feels very sad, anxious and without hope and often has physical symptoms such as being unable to sleep, etc*".[1] **For example**:"*Debts and suicidal thoughts. During the year of 2021 I've Lost* 3 *relatives and got a debt of* 2000 *USD due to that. I'm getting more and more depressed and i'm not sleeping well...*".

- *Suicide*: It is defined as "*a death that happens when someone harms themselves because they want to end their life. It is one of the mental health problems and a leading cause of death.*"[2] **For example**: "Being broke makes me want to kill myself Bills keep coming.

[1] https://bit.ly/3LJ3R16
[2] https://medlineplus.gov/suicide.html

I don't have enough money for anything. Saving change just to eat...".

- *Stress*: It is defined as "*any type of change that causes physical, emotional, or psychological strain*".[3] **For example**: "I'm having bad thoughts...I'm about to lose my house and no where to go. I'm 4 months behind on my mortgage (got laid off from my job because of covid)...".

3.2 Dataset Preparation

This sub-section presents the dataset collection and pre-processing, and filtration of candidate mental health financial text from these datasets.

3.2.1 Dataset Collection and Pre-processing

In this study, `Reddit` posts are taken to prepare the dataset. We have used `Reddit` API[4] using `PRAW`[5] wrapper to retrieve posts based on mental health and non-mental health. All posts are collected via *subreddits*, which presents specific topic on `Reddit` and preceded by r/. We have used *subreddits* based on mental health, such as r/mentalhealth, r/mmfb (make me feel better), r/offmychest, r/traumatoolbox, r/anxietyhelp, r/CPTSD (complex post traumatic stress disorder), r/depression, and r/SuicideWatch to prepare mental health datasets. Likewise, *subreddits*, such as r/happy, r/mademesmile, r/makemesmile, r/financeadvice, r/creditcards, r/wholesome, r/economics, r/financialindependence, r/financialplanning, r/investing, r/personalfinance, r/pftools, and r/tax are considered to prepare a non-mental health datasets. Thus, 320000 mental health instances and 270000 non-mental health instances are finalized. Further details about the datasets are given in the sub-section 5.1.

All collected raw datasets consist of many noise and non-literal information. Removal of such unwanted information is crucial to achieve good classification accuracy. Considering this, we have performed data cleaning steps and removed symbols, hexa characters, mentions, hashtags, slash, exclamation, quotation, and punctuation marks, unwanted links, ampersands, extra dots and spaces, digits, and non-ASCII characters. In the end, we have converted raw text into the lower-case form.

3.2.2 Dataset Filtration

On analyzing all mental health datasets, it is found that all mental health instances do not contain financial information. Also, there is no availability of benchmark datasets on financial mental health. Considering this, two-steps filtering technique is applied to filter only those mental health instances from the datasets, which are context- and semantic-wise related to finance.

To this end, in step-1, we have compiled a list of keywords based on unigram and bigram tokens from (Guan et al., 2022) work along with our generated keywords (i.e, unigram and bigram) tokens related to finance or financial situation, as given in table 1. We filter candidate mental health instances using regular expression-based criteria, in which if at least one of the keywords (unigram/bigram) matches with pre-processed mental health instance from table 1.

Table 1: A List of keywords based on finance or financial situation

List of keywords (unigram and bigram)
'income', 'debt', 'loan', 'mortgage', 'finance', 'economy', 'job', 'broke', 'poor', 'poverty', 'homeless', 'salary', 'money', 'bank', 'savings', 'scam', 'robbery', 'deprivation', 'loss', 'fund', 'earn'', 'payroll'', 'earning', 'wage', 'fired', 'livelihood', 'compensation', 'revenue', 'allowance', 'payoff', 'wealth', 'asset', 'economic situation', 'economic stat', 'economic condition', 'economic position', 'economic hardship', 'economic str', 'economic difficult', 'financial situation', 'financial position', 'financial condition', 'financial stat', 'financial hardship', financial difficult'

Thereafter, in step-2, we have applied zero-shot learning (ZSL), which aims to perform predictions without having seen labelled training instances. It is widely used in NLP and text classification problems, in which combination of seen and unseen labelled via auxiliary information are taken to encode the available discrimination attributes of an instances (Xian et al., 2018). `Facebook/Bart-large-mnli`[6] is one of the popular ZSL method which is based on `Hugging`

[3] `https://bit.ly/2yYHnVu`
[4] `https://www.reddit.com/dev/api/`
[5] `https://bit.ly/3r7690r`

[6] `https://huggingface.co/facebook/bart-large-mnli`

Face-based Transformer model trained on multi natural language interface (*aka*, MNLI) dataset. It is based on two concepts – *premises* (refers to the instances which is to be classified) and *hypothesis* (refers to the number of class labels) (Plaza-Del-Arco et al., 2022). A confidence score is generated for each *hypothesis* of a given *premise*, and based on the highest confidence score, the *premise* is predicted. In this study, we set threshold of 0.89 for the highest confidence score.

Considering this, each filtered mental health instance in step-1 is passed as a *premise* and accordingly two *hypothesis* are taken - *financial* and *non-financial* for labelling purpose. As a result, candidate financial mental health instances are collected based on the highest confidence score. Figure 2 presents an example based on the *premise* and *hypothesis* concepts related to MNLI to ZSL methods. In this example, the representation of finance related mental health is *entailed*. Hence, it is predicted as MFH instance.

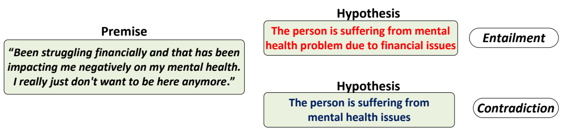


Figure 2: An example for data filtration using MNLI to ZSL methods for mental health instance.

Figure 3 presents a better visual insights and effectiveness of data filtration step on collected Reddit posts via two word clouds. It can be seen from this figure that only frequent words related to mental health are seen before data filtration step, whereas top-frequent words related to finance and mental health are found after data filtration step.

(a) Before data filtration (b) After data filtration

Figure 3: Effect of data filtration step via word cloud

4 Proposed Model

This section presents a detailed description of the proposed IMFinE model. It includes an input layer, two embeddings (MentalBERT and FinBERT), CNN, BiGRU, followed by a dense and output layers. Figure 4 presents the architecture of the proposed IMFinE model.

4.1 Input Layer

Given the candidate MHF or NMHF as input text n consisting of w_n words, the input layer tokenizes each word available in its textual contents. Each token is indexed in a dictionary and converted into a numeric vector v such that it replaces according to its index value as per the dictionary. The length of v is different because of the varying input length size. Hence, a fixed-length l of input vectors is maintained for each input text. Thus, v is transformed to a padded-vector p of 200 fixed-length, such that $|p| = l \geq |v|$. The fixed-length resulting vector $p \in R^{1\times l}$ is then forwarded to the embedding layer.

4.2 Embedding Layers

In the last few years, BERT has gained immense popularity because it deals with contextual information in both directions via Transformer. MentalBERT is a pre-trained masked language model using BERT for mental health detection tasks, whereas FinBERT is a pre-trained NLP model and it is built by training the BERT model in the financial domain over huge financial data. Considering this, we leverage these two specific BERT models – MentalBERT and FinBERT based on our problem statement to retrieve contextualized language representations based on mental health and finance from the input text.

The generated input vector from the input layer is passed to both parallel embedding layers with a maximum sequence length of 200. We consider 768-dimensional word vector representation for both MentalBERT and FinBERT embeddings. Consequently, the relevant contextual information based on mental health and finance is retrieved from both embeddings in parallel. The encoded representation from two embeddings is concatenated and that gives a rich comprehensive contextual representation of the input text. Further, it is passed to the CNN layer.

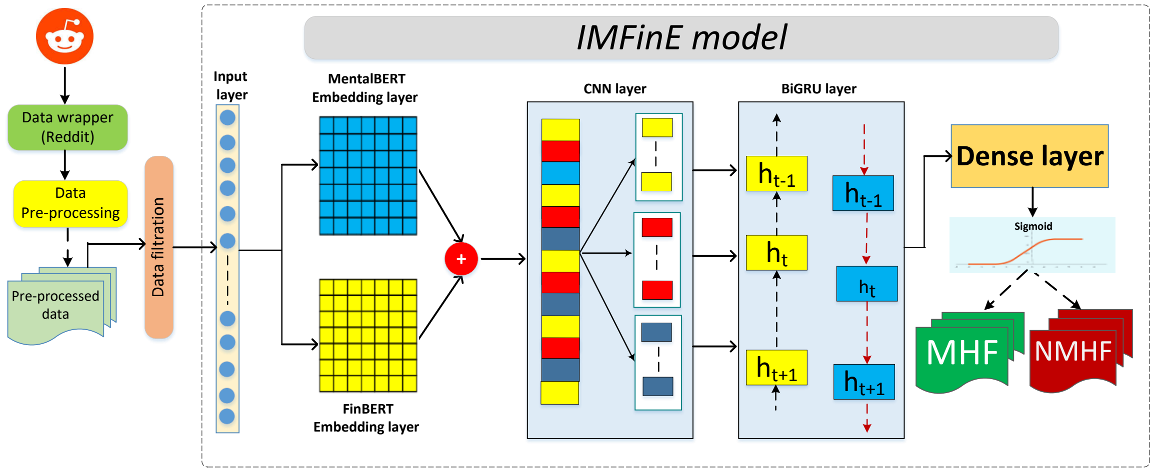


Figure 4: Architecture of the proposed `IMFinE` model

4.3 CNN Layer

Kim (2014) proposed CNN which extracts low-level semantic and syntactic features automatically. It extracts contextual local and positional invariant features, performs several convolution operations on the received input texts, and generates the global feature vector. In this study, we have used 64 filters of window size 3 that moves the comprehensive embedding vectors for extracting features related to mental health and finance. Max pooling operation of pool size 3, `ReLU` activation function, and 64 filters perform the convolutional operation from top to bottom, and extract the feature sequence as f_t = [f_1, f_2, ..., f_{64}] based on mental health and finance for the input text. The n^{th} feature sequence, f_t from word window x_t is generated, as given in 1. Finally, the filter outputs are concatenated to give the resultant mental health and finance-based feature representation, which is forwarded to the BiGRU layer.

$$f_t = r(w_t \cdot x_t + b) \tag{1}$$

4.4 BiGRU Layer

Liu and Guo (2019) proposed GRU, which deals with long-term temporal dependencies without dealing the vanishing gradient descent problem. It consists of two gates and operational in both forward and backward directions as forward GRU and backward GRU, respectively. In this study, the role of the forward GRU and backward GRU is to generate succeeding feature sequences (f_t to f_{64}) and preceding feature sequences (f_{64} to f_t), respectively for mental health and finance related latent contextual feature sequences from the extracted features of the CNN layer. Thus, equations 2 and 3 show forward and backward directions of BiGRU outcomes, respectively. The resultant combined outcome of both forward and backward directions is passed to the dense layer.

$$\overrightarrow{gru_f} = \overrightarrow{GRU}(L_{f_t}), n \in [1, 64] \tag{2}$$

$$\overleftarrow{gru_b} = \overleftarrow{GRU}(L_{f_t}), n \in [64, 1] \tag{3}$$

4.5 Dense and Output Layers

In this study, both datasets are divided as a training set, a testing set, and a validation set, wherein 70% is used for training, 20% is used for testing, and 10% is used for validation, we use 30 epochs with early stopping and `Adam` as an optimization algorithm. The fully connected dense layer gives features set based on the outcome of previous layers divisible into two classes. Finally, the *sigmoid* activation function is used upon the dense layer, and *binary cross-entropy* loss function is used for classifier training which interprets input text labelled as MHF or NMHF.

5 Experimental Setup and Results

This section presents the experimental details of the proposed `IMFinE` model. It includes the statistical details of the datasets, experimental and hyper-parameter settings, evaluation metrics, followed by evaluation results and comparative analysis, and ablation study.

5.1 Datasets

This section presents the datasets used in this study. We have considered three datasets from `Reddit`, out of which two are benchmark datasets. A short description of the two benchmark datasets is given below:

- **Dreddit dataset**: Turcan and McKeown (2019) proposed this dataset, where they collected 1857 *stress* and 1696 *non-stress* instances to prepare training and testing datasets.
- **SDCNL dataset**: Haque et al. (2021) proposed this dataset, where they collected 915 *depression* and 981 *suicide* instances to prepare training and testing datasets.

We have combined *stress*, *depression*, and *suicide* instances of both benchmark datasets for mental health category and *non-stress* instances as non-mental health category. As a result, a total number of 3753 and 1696 instances is finalized for mental health and non-mental health categories, respectively after combining both benchmark datasets and named it as DS-1. Besides that, we have prepared one dataset using `Reddit`, as discussed in sub-section 3.2.1, and named it as DS-2. Further, we have applied dataset filtration techniques across all datasets as discussed in the sub-section 3.2.2. Table 2 presents the final statistics of the datasets after data filtration steps.

Table 2: Final statistics of the datasets after data filtration steps

Datasets ↓	**MHF**	**NMHF**	**Total**
DS-1	509	2613	3122
DS-2	8400	8400	16800

5.2 Experimental and Hyperparameter Settings

This section presents the details about the experimental and hyperparameter settings used for the implementation of the proposed `IMFinE` model. Table 3 presents the summary of the experimental settings. Likewise, table 4 presents the summary of the hyperparameter settings.

Table 3: A summary of the experimental settings

Environment	**Configurations**
Machine	`Intel Haswell`
Operating system	`Ubuntu-20.04`
Memory (RAM)	32 GB
GPU	`NVIDIA Tesla`
Language	`Python 3.9`
Neural network library	`Keras 2.10.0`
`Reddit` API wrapper	`PRAW`

Table 4: A summary of the hyperparameter settings

Hyperparameter	**Values**
Batchsize	32
Padding	200
Spatial dropout	0.2
Dropout	0.2
#neurons (GRU)	100
#filters (CNN)	64
window size (CNN)	3
pool size (CNN)	3

5.3 Evaluation Metrics

The proposed model is evaluated using four standard evaluation metrics – *precision*, *recall*, *f-score*, and *accuracy*. Equations 4 to 7 define these metrics formally using the terms – *true positive* (*TP*), *false positive* (*FP*), *true negative* (TN), and *false negative* (FN). TP represents the total number of correctly classified MHF instances. FP represents the total number of incorrectly classified MHF instances. TN represents the total number of correctly classified NMHF instances. Finally, FN represents the total number of incorrectly classified NMHF instances.

$$Precision\ (P) = \frac{TP}{TP + FP} \tag{4}$$

$$Recall\ (R) = \frac{TP}{TP + FN} \tag{5}$$

$$F\text{-}score\ (F) = \frac{2 \times P \times R}{P + R} \tag{6}$$

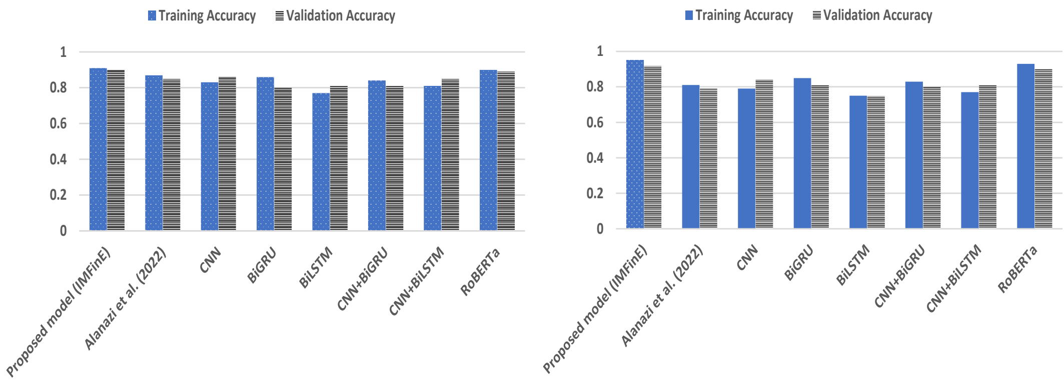
$$Accuracy\ (A) = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

5.4 Evaluation Results and Comparative Analysis

In this section, we present the evaluation results of the proposed `IMFinE` model and compared it with a recent study proposed by Alanazi et al. (2022), wherein authors addressed

Table 5: Performance results on both datasets

Datasets →	DS-1			DS-2		
Methods ↓	**P**	**R**	**F**	**P**	**R**	**F**
Proposed model (`IMFinE`)	**0.82**	**0.86**	**0.84**	**0.95**	**0.93**	**0.94**
Alanazi et al. (2022)	0.74	0.80	0.77	0.91	0.92	0.91
CNN	0.72	0.83	0.77	0.90	0.89	0.89
BiGRU	0.80	0.84	0.82	0.93	0.91	0.92
BiLSTM	0.76	0.83	0.79	0.91	0.91	0.91
CNN+BiGRU	0.70	0.84	0.76	0.90	0.88	0.89
CNN+BiLSTM	0.67	0.81	0.73	0.88	0.90	0.89
`RoBERTa`	0.81	0.85	0.83	0.94	0.85	0.89



(a) DS-1 dataset

(b) DS-2 dataset

Figure 5: Training versus validation accuracy on DS-1 and DS-2 datasets

mental health monitoring using sentimental analysis of financial text. We have also compared our proposed `IMFinE` model with neural network-based baseline methods. Table 5 presents the performance evaluation results on both datasets. It can be observed that proposed `IMFinE` model shows good result on both datasets. It receives *f-score* of 0.84 and 0.94 for DS-1 and DS-2, respectively. This refers to an interesting observation that the proposed model is performing better on both unbalanced (DS-1) and balanced (DS-2) datasets. It can also be observed that it performs significantly better than SOTA work and neural network baseline methods in terms of *precision*, *recall*, and *f-score*. However, BiGRU shows the best performance across neural network baseline models, but our proposed model is performing more better and leveraging contextual information from both embedding on both datasets and receiving latent contextual information sequences. One more interesting observation is the proposed `IMFinE` model performs better with smaller dataset (i.e., DS-1) as well.

Figure 5 presents the visualization of the training versus validation accuracy on both datasets. It shows that the proposed model performs impressive in terms of training versus validation accuracy on DS-1 and DS-2 datasets. It can also be seen that the proposed model shows significantly better results in comparison to the SOTA and neural network baseline results.

Table 6: Ablation study of the proposed model in terms of *f-score* on DS-1 and DS-2 datasets. datasets.

Proposed model (`IMFine`)	DS-1	DS-1
All layers	**0.84**	**0.94**
Without CNN	0.80	0.91
Without BiGRU	0.81	0.90

5.5 Ablation Study

This section presents the ablation study of the proposed model to show component-wise analysis.

Table 6 shows the proposed model with all layers performs better. The performance of the model is effected without considering CNN and BiGRU.

6 Conclusion

We have proposed a new problem for mental health detection in financial context. It has started with two-level data filtration steps to filter candidate mental health instances in financial domain. We have proposed a novel `IMFinE` model which is consisted of input, `BERT`-based relavant contextual embeddings (`MentalBERT` and `FinBERT`), CNN, BiGRU, and, output followed dense layers. The proposed `IMFinE` model has evaluated over `Reddit` datasets and experimental results have shown significantly better results in comparison to SOTA and several neural network baseline methods. The evaluation of the `IMFinE` model in a multi-modal setting could be an important direction of research. `IMFinE` model can also be extended over multi-lingual datasets.

References

Muhammad Abulaish and Ashraf Kamal. 2018. Self-deprecating sarcasm detection: An amalgamation of rule-based and machine learning approach. In *Proceeding IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 574–579. IEEE.

Muhammad Abulaish, Ashraf Kamal, and Mohammed J. Zaki. 2020. A survey of figurative language and its computational detection in online social networks. *ACM Transactions on the Web (TWEB)*, 14(1):1–52.

Saad A. Alanazi, Ayesha Khaliq, Fahad Ahmad, Nasser Alshammari, Iftikhar Hussain, Muhammad A. Zia, Madallah Alruwaili, Alanazi Rayan, Ahmed Alsayat, and Salman Afsar. 2022. Public's mental health monitoring via sentimental analysis of financial text using machine learning techniques. *International Journal of Environmental Research and Public Health*, 19(15):9695.

Iqra Ameer, Muhammad Arif, Grigori Sidorov, Helena Gòmez-Adorno, and Alexander Gelbukh. 2022. Mental illness classification on social media texts using deep learning and transfer learning. *arXiv preprint arXiv:2207.01012*.

Gunjan Ansari, Muskan Garg, and Chandni Saxena. 2021. Data augmentation for mental health classification on social media. In *Proceedings of the Sixteen International Conference on Natural Language Processing*, pages 152–161, National Institute of Technology, Silchar, India. NLP Association of India.

Dogu Araci. 2019. Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*.

Munmun D. Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In *Proceeding the Seventh International AAAI Conference on Weblogs and Social Media*.

Glen Coppersmith, Kim Ngo, Ryan Leary, and Anthony Wood. 2016. Exploratory analysis of social media prior to a suicide attempt. In *Proceedings of the third workshop on computational linguistics and clinical psychology*, pages 106–117.

Chris Fitch, Robert Chaplin, Colin Trend, and Sharon Collard. 2007. Debt and mental health: the role of psychiatrists. *Advances in Psychiatric Treatment*, 13(3):194–202.

Giorgos Giannakakis, Matthew Pediaditis, Dimitris Manousos, Eleni Kazantzaki, Franco Chiarugi, Panagiotis G. Simos, Kostas Marias, and Manolis Tsiknakis. 2017. Stress and anxiety detection using facial cues from videos. *Biomedical Signal Processing and Control*, 31:89–101.

Naijie Guan, Alessandra Guariglia, Patrick Moore, Fangzhou Xu, and Hareth Al-Janabi. 2022. Financial stress and depression in adults: A systematic review. *PloS one*, 17(2):e0264041.

Ayaan Haque, Viraaj Reddi, and Tyler Giallanza. 2021. Deep learning for suicide and depression identification with unsupervised label correction. In *Proceedings of the Thirtieth International Conference on Artificial Neural Networks, Bratislava, Slovakia, 2021*, pages 436–447.

Yeen Huang and Ning Zhao. 2020. Generalized anxiety disorder, depressive symptoms and sleep quality during covid-19 outbreak in china: A web-based cross-sectional survey. *Psychiatry research*, 288:112954.

Shaoxiong Ji, Tianlin Zhang, Luna Ansari, Jie Fu, Prayag Tiwari, and Erik Cambria. 2021. Mentalbert: Publicly available pretrained language models for mental healthcare. *arXiv preprint arXiv:2110.15621*.

Ashraf Kamal and Muhammad Abulaish. 2019a. An lstm-based deep learning approach for detecting self-deprecating sarcasm in textual data. In *Proceedings of the 18th International Conference on Natural Language Processing*, pages 201–210, International Institute of Information Technology, Hyderabad, India. NLP Association of India.

Ashraf Kamal and Muhammad Abulaish. 2019b. Self-deprecating humor detection: A machine learning approach. In *International Conference of the Pacific Association for Computational Linguistics*, pages 483–494. Springer.

Ashraf Kamal and Muhammad Abulaish. 2022. Cat-bigru: Convolution and attention with bi-directional gated recurrent unit for self-deprecating sarcasm detection. *Cognitive Computation*, 14(1):91–109.

Shakir Khan, Mohd Fazil, Vineet K. Sejwal, Mohammed A. Alshara, Reemiah M. Alotaibi, Ashraf Kamal, and Abdul R. Baig. 2022a. Bichat: Bilstm with deep cnn and hierarchical attention for hate speech detection. *Journal of King Saud University-Computer and Information Sciences*, 34(7):4335–4344.

Shakir Khan, Ashraf Kamal, Mohd Fazil, Mohammed A. Alshara, Vineet K. Sejwal, Reemiah M. Alotaibi, Abdul R. Baig, and Salihah Alqahtani. 2022b. Hcovbi-caps: Hate speech detection using convolutional and bi-directional gated recurrent unit with capsule network. *IEEE Access*, 10:7881–7894.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar. Association for Computational Linguistics.

Russell Li and Zhandong Liu. 2020. Stress detection using deep neural networks. *BMC Medical Informatics and Decision Making*, 20(11):1–10.

Huijie Lin, Jia Jia, Liqiang Nie, Guangyao Shen, and Tat-Seng Chua. 2016. What does social media say about your stress?. In *IJCAI*, pages 3775–3781.

Gang Liu and Jiabao Guo. 2019. Bidirectional lstm with attention mechanism and convolutional layer for text classification. *Neurocomputing*, 337:325–338.

Minsu Park, Chiyoung Cha, and Meeyoung Cha. 2012. Depressive moods of users portrayed in twitter. In *Proceedings of the Eighteen ACM International Conference on Knowledge Discovery and Data Mining, SIGKDD 2012*, pages 1–8.

Flor Miriam Plaza-Del-Arco, María-Teresa Martín-Valdivia, and Romann Klinger. 2022. Natural language inference prompts for zero-shot emotion classification in text across corpora. *arXiv preprint arXiv:2209.06701*.

Aryan Rastogi, Liu Qian, and Erik Cambria. 2022. Stress detection from social media articles: New dataset benchmark and analytical study. In *Proceeding of the IEEE World Conference of Computational Intelligence*.

XU Ronghua and Zhang Qingpeng. 2016. Understanding online health groups for depression: Social network and linguistic perspectives. *Journal of medical Internet research*, 18(3):e5042.

Ramit Sawhney, Harshit Joshi, Lucie Flek, and Rajiv Shah. 2021. Phase: Learning emotional phase-aware representations for suicide ideation detection on social media. In *Proceedings of the 16th conference of the european chapter of the association for computational linguistics: main volume*, pages 2415–2428.

Guangyao Shen, Jia Jia, Liqiang Nie, Fuli Feng, Cunjun Zhang, Tianrui Hu, Tat-Seng Chua, and Wenwu Zhu. 2017. Depression detection via harvesting social media: A multimodal dictionary learning solution. In *IJCAI*, pages 3838–3844.

Mike Thelwall. 2017. Tensistrength: Stress and relaxation magnitude detection for social media texts. *Information Processing & Management*, 53(1):106–121.

Sho Tsugawa, Yusuke Kikuchi, Fumio Kishino, Kosuke Nakajima, Yuichi Itoh, and Hiroyuki Ohsaki. 2015. Recognizing depression from twitter activity. In *Proceedings of the Thirty Third annual ACM conference on human factors in computing systems*, pages 3187–3196.

Elsbeth Turcan and Kathleen McKeown. 2019. Dreaddit: A reddit dataset for stress analysis in social media. *EMNLP-IJCNLP 2019*, page 97.

Gali H Weissberger, Laura Mosqueda, Annie L Nguyen, Anya Samek, Patricia A Boyle, Nguyen Caroline P, and S Duke Han. 2020. Physical and mental health correlates of perceived financial exploitation in older adults: Preliminary findings from the finance, cognition, and health in elders study (finches). *Aging & mental health*, 24(5):740–746.

Yongqin Xian, Christoph H. Lampert, Bernt Schiele, and Zeynep Akata. 2018. Zero-shot learning—a comprehensive evaluation of the good, the bad and the ugly. *IEEE transactions on pattern analysis and machine intelligence*, 41(9):2251–2265.

Yuanyuan Xue, Qi Li, Li Jin, Ling Feng, David A. Clifton, and Gari D. Clifford. 2014. Detecting adolescent psychological pressures from micro-blog. In *International Conference on Health Information Science*, pages 83–94. Springer.