D-HAN-Net: A Hybrid Dual-Stream Architecture for Corporate Bankruptcy Prediction via Multimodal Fusion

Md. Abul Kalam Azad[1], Abdul Kadar Muhammad Masum[2*], Najmus Saadat[3], Esrat Jahan[4],
Ramona Birău[5], Virgil Popescu[6], Iuliana Carmen Bărbăcioru[7], Ștefan Mărgăritescu[8]
Dept. of Business and Technology Management, Islamic University of Technology, Gazipur 1704, Bangladesh[1]
Dept. of Computer Science and Engineering, Southeast University, Dhaka 1208, Bangladesh[2]
Computational Intelligence Lab, Southeast University, Dhaka 1207, Bangladesh[3,4]
"Eugeniu Carada" Doctoral School of Economic Sciences, University of Craiova, Craiova, Romania[5,8]
Faculty of Economic Science, Constantin Brâncusi University of Târgu Jiu, 210185 Târgu Jiu, Romania[5]
Faculty of Economics and Business Administration, University of Craiova, Craiova, Romania[6]
Faculty of Engineering, University of Târgu Jiu, Gorj county, Romania[7]

***Abstract*—Early detection of corporate bankruptcy is essential for maintaining financial market stability and reducing systemic risk. However, existing predictive models often struggle under conditions of extreme class imbalance. Traditional approaches either analyze financial ratios or textual disclosures in isolation, while conventional cross-modal fusion strategies tend to dilute rare distress signals when integrating structured accounting metrics with qualitative management narratives. Consequently, subtle indicators of impending corporate failure are frequently overshadowed by dominant non-distress patterns, limiting the effectiveness of existing predictive systems. To bridge this methodological limitation, this study proposes D-HAN-Net, a hybrid dual-stream deep learning architecture. This framework is intended to address minority class suppression by dynamically balancing structured financial ratios with unstructured textual disclosures. Specifically, the model processes numerical indicators through a gated residual network to capture complex patterns. Simultaneously, it extracts semantic cues from corporate reports via a FinBERT-driven bidirectional GRU. These dual modalities are then aligned using a learnable cross-modal attention fusion gate, jointly optimized with focal loss. Experimental evaluations on a comprehensive multimodal dataset, utilizing stratified splits demonstrate that D-HAN-Net significantly outperforms state-of-the-art baselines, achieving a predictive accuracy of 94.00%, an F1-score of 88.00%, and an AUC of 0.9734. Practically, this framework equips investors, financial institutions, and regulatory authorities with a decisive early warning system. It enables proactive risk management by detecting subtle distress signals before corporate failure becomes irreversible. Furthermore, extensive stability testing and ablation analysis confirm that the model's superior predictive reliability is highly robust against sampling uncertainty, fundamentally relying on the synergistic integration of all its architectural modules.**

Keywords*—Bankruptcy prediction; multimodal deep learning; corporate disclosures; cross-modal fusion; class imbalance; early warning system

I. Introduction

Corporate bankruptcy prediction has become a central research problem in financial analytics, as the early identification of financially distressed firms is very critical for investors, creditors, regulators and other market participants. Accurate prediction models can minimize the potential losses, facilitate the timely interventions, and enhance the distribution of financial resources across the capital markets. The number of corporate bankruptcies is also generally believed to be a signal of overall economic health and financial stability, with recent studies like Ji et al. [1] emphasizing that such models serve as critical early warning systems for stakeholders. For this reason, bankruptcy prediction is still an important part of financial risk assessment and corporate decision support. Traditionally, the goal of bankruptcy prediction models has been to assess the financial situation of a firm and forecast the probability of failure before distress is irreversible, a concept further advanced by Lin et al. [2] through modern explainable machine learning integrations. Most previous research tackled this task by means of the financial statement analysis by using financial statements accounting based indicators created from balance sheets, income statements and cash flow statements. These structured numerical variables have long been used as the basis for both statistical and machine learning based bankruptcy prediction models. Prior work indicates that machine learning can surpass traditional statistical methods by capturing complex nonlinear relationships between financial indicators. Recent explainable and hybrid approaches proposed by Fasano et al. [32] and Ainan et al. [34] further demonstrate the growing effectiveness of intelligent predictive systems in bankruptcy analysis. For instance, Liu et al. [3] report that machine learning models optimized with advanced algorithms are able to perform prediction much better, and Jiang et al. [4] highlight the growing importance of such intelligent computational approaches to financial distress studies. However, despite this progress, bankruptcy prediction remains challenging due to the rarity, heterogeneity, and complexity of real-world corporate failure patterns. Corporate bankruptcy prediction can be viewed as an imbalanced multimodal financial risk classification problem in which structured accounting variables and unstructured narrative disclosures must be jointly modeled.

A key issue in this literature is the extreme class imbalance that is present in most bankruptcy data. In practice, bankrupt companies are a small fraction of all companies, which often

*Corresponding author

leads prediction models to be biased towards the majority class of healthy companies and perform poorly on the minority class of distressed companies. Lombardo et al. [5] point out the practical difficulty of bankruptcy prediction under highly imbalanced settings, whereas Ranjbaran et al. [6] demonstrate that augmentation based strategies might lead to better minority detection, but at the same time can lead to instability and increased false positive rates. Beyond the imbalance problem, another important limitation is the high level of dependence of many prediction systems on structured financial ratios alone. Although such variables capture quantitative aspects of profitability, liquidity, leverage and solvency, the development of financial distress rarely occurs through numerical deterioration only. In real corporate situations, the warning signs of distress may often appear in managerially talk, auditor's concerns and forward looking uncertainty before it may be fully reflected in the accounting results. Corporate narrative disclosures, particularly the Management Discussion and Analysis section and auditor reports, therefore, lend a further dimension of risk relevant information. Mai et al.[7] demonstrate the use of textual disclosures from corporate filings to further improve bankruptcy prediction since they capture the semantic patterns that are unavailable in structured data. Related accounting research also lends credence to the informational value of narrative reporting. Mayew et al. [8] find that Management Discussion and Analysis disclosures to deliver incremental evidence about the firm's ability to continue as a going concern. Li et al. [9] show that readability in financial reporting is associated with earnings persistence and transparency while Loughran and McDonald [10] demonstrate that linguistic tone in corporate filings carries risk related information beyond conventional accounting measures. From a theoretical perspective these results imply that numerical indicators and textual disclosures are complementary as opposed to competing information channels.Financial ratios summarize realized economic outcomes, while narrative disclosures communicate managerial interpretation, uncertainty and emerging risks that may not yet be fully reflected in accounting variables. This view is consistent with both information asymmetry as well as signaling theories of financial reporting, where managers convey firm conditions in the form of both formal accounting results and qualitative disclosures. Yet most bankruptcy prediction studies still consider these two modalities separately. Although recent multimodal frameworks such as Mancisidor and Aas [37] indicate that combining textual and numerical information can substantially improve predictive capability. As a result, existing approaches often fail to effectively capture cross-modal interactions between quantitative financial deterioration and qualitative narrative distress signals.

This limitation leads to an obvious research gap and directly leads in to the present study. Despite the recent progress in deep learning, specifically transformer-based language models like BERT by Devlin et al. [11], the representation of long textual documents has been significantly improved. However, bankruptcy prediction research is still missing a unified framework that can effectively integrate structured financial indicators with corporate narrative disclosures under highly imbalanced conditions. Recent evidence from Chen et al. [12] suggests that textual information extracted from the annual reports can improve the predictive performance when integrated into learning-based systems. suggests that textual information extracted from annual reports can improve predictive performance when integrated into learning based systems. Similar findings have also been reported in recent multimodal and hybrid bankruptcy prediction studies such as Arno et al. [36], but the broader challenge of theoretically and computationally combining text and numerical signals is still to be solved. Nevertheless, modeling lengthy financial disclosures remains challenging because important distress-related evidence may be distributed across multiple document segments rather than localized within short textual spans. This encourages the following research question: How can structured financial indicators and narrative corporate disclosures be successfully combined in a unified predictive model in order to better detect corporate financial distress under highly imbalanced bankruptcy datasets?

To answer this question, this study proposes D-HAN-Net, a hybrid dual-stream deep learning architecture that considers the analysis of structured numerical indicators and corporate textual disclosures jointly. The framework uses a gated residual network to process financial variables to model nonlinear dependencies among ratios. It also employs a transformer-based language encoder followed by a hierarchical recurrent aggregation layer to model the text disclosures and retain the contextual semantic structure across long documents. This design also improves computational efficiency and training stability under limited bankruptcy data conditions. A cross-modal fusion mechanism then combines the two modalities and dynamically balances the relative contributions of the two modalities during the prediction process. The main contributions of this study are summarized as follows: First, it constructs a multimodal deep learning architecture for joint modeling of financial ratios and corporate narrative disclosures for bankruptcy prediction. Second, it presents a cross-modal fusion strategy that leverages the complementary nature between numerical and textual signals in an integrated architecture. Third, it illustrates the efficacy and robustness of the proposed framework with extensive empirical analysis of the framework, including baseline comparison, ablation analysis, and cross seed stability analysis.

II. Literature Review

Bankruptcy prediction has long been studied in financial economics, accounting, and data science. Over the years, the methods of research have evolved from old statistical models to new machine learning and deep learning techniques. Financial distress theory and information asymmetry literature imply that corporate distress signals arise from various sources, which includes financial statements and corporate disclosures. Consequently, both financial indicators and narrative information can offer valuable signals for identifying financial difficulty faced by firms in the form of both quantitative financial indicators and qualitative narrative information. This section provides an overview of the theoretical background of bankruptcy prediction, and an overview of previous research works concerning machine learning models, text financial analysis and imbalance aware learning models.

Corporate bankruptcy generally has been explained in terms of theories of financial distress between firm failure and deterioration in financial structure, liquidity, increasing leverage, and falling profitability. Bankruptcy risk, therefore, becomes visible through signals such as the worsening of liquidity, unstable cash flows, declining profitability and growing pressure to ensure solvency. For this reason, financial indicators based on financial statements have been the main inputs for bankruptcy prediction models for a long time [13]. However, financial distress is not only reflected in the accounting numbers. Disclosure theory implies that managers are communicating risk via narrative disclosures in annual reports and regulatory filings. When firms find themselves under financial stress, discussions among managers usually involve problems with operations, problems with financing, or restructuring activities. Such signals often occur in advance of the time when deterioration is fully reflected in the accounting ratios. Financial language models like FinBERT can capture these contextual cues because they are trained on financial corpora and are able to spot domain specific expressions related to uncertainty and financial instability. Taken together, these views mean that financial distress is multidimensional. Numerical indicators provide information on measured deterioration in firm performance, while narrative disclosures provide managerial perceptions and advance warning signals about financial stress. Integrating financial indicators and FinBERT derived semantic representations, therefore, offers a theoretically grounded basis for enhancing bankruptcy prediction models [14], [15].

Machine learning techniques have been popularly used to predict bankruptcy because of their capability of capturing complex nonlinear relationships between financial variables and bankruptcy outcomes. Studies have been done on the neural network and the sequential learning models to detect the financial distress pattern. Kim et al. showed that the modeling of temporal dependencies between financial indicators is useful for enhancing prediction performance [16]. Hosaka proposed transforming financial ratios into image representations and using convolutional neural networks to identify patterns in structure of financial data [17]. Pellegrino et al. further developed an architecture of multiple heads Long Short Term Memory to learn temporal dynamics from accounting time series data [18]. Ensemble learning methods have also attracted a lot of attention. Recent interpretable ensemble frameworks combining models like LightGBM and CatBoost with SHAP analysis demonstrate significantly more robust and transparent predictions across diverse market conditions [19]. Liang et al. showed that the integration of corporate governance index and financial index can improve the predictive effect and will show the importance of diverse corporate characteristics in the predictive model [20].

Recent researches have been devoted to advanced models that can be applied to high-dimensional financial dataset. Gradient boosting methods like LightGBM have been found to have good predictive power in financial risk modeling [21]. Ensemble approaches have also been shown to have great potential. Entropy based feature transformation strategies which enhances predictive performance in financial risk modeling was introduced by Carta et al. [22], [23]. Stacking based ensemble methods further suggests that a combination of classifiers will yield more stable and reliable results than relying on a single model [24]. Feature selection methods have also been proven to be efficient to identify the most informative indicators to predict bankruptcy [25]. Research and surveys have shown that machine learning and deep learning methods typically outperform conventional statistical models in bankruptcy prediction tasks [26], [27]. Financial distress is communicated not only through numerical ratios but also through managerial narratives in corporate disclosures. Managers often adjust the tone and wording of reports when firms face liquidity pressure, operational challenges, or solvency concerns. As a result, textual disclosures may reveal early signals of distress before these conditions become fully visible in financial statements. Transformer based language models such as BERT can capture contextual semantic relationships in textual data [11]. In financial applications, domain-adapted models such as FinBERT, are particularly effective at identifying linguistic patterns related to uncertainty, litigation risk, and financial instability. Chen et al. demonstrated that incorporating textual features from annual reports significantly improves bankruptcy prediction performance [12]. Lombardo et al. further showed that deep language models can effectively analyze complex corporate disclosure documents [28].

One of the biggest challenges in predicting bankruptcy lies in the extreme class imbalance with financial datasets, with bankrupt firms only making up a small percentage of observations. This imbalance will often lead to predictive models giving preference to the majority class, and making it difficult for them to accurately determine correct cases of bankruptcy. To overcome this problem, imbalance aware learning strategies have been proposed. Recently, Papík demonstrated the significant impact of advanced data resampling approaches, such as SMOTE and random oversampling, on improving minority class detection for imbalanced financial datasets [29]. Kubat and Matwin have introduced one sided selection techniques that reduce the imbalance by eliminating redundant majority class samples from the training data [30]. These approaches have become important parts of modern financial risk predicting systems. Despite some significant progress, there are still some limitations in the literature. Many of the researchers mainly focus on financial indicators and pay little attention to qualitative signals in corporate disclosures. Studies that look at the analysis of textual information tend to look at the narrative signals independent of financial variables leaving the interaction between these two parameters insufficiently explored. Furthermore, many models work such that they fit within a single modality framework and cannot, therefore, capture complementary distress signals from heterogeneous data sources.

To overcome these limitations, this research study proposes D-HAN-Net, a dual stream hybrid deep learning architecture that jointly analyzes structured financial data and semantic information from the corporate disclosures. The model uses separate encoders for numerical and textual modalities and uses an adaptive cross-modal fusion mechanism to dynamically balance contributions of the two modalities during the prediction. By incorporating heterogeneous financial signals into an integrated framework, the proposed approach is expected to enhance the bankruptcy prediction performance and result in a more robust bankruptcy early warning system under class imbalanced conditions.

III. METHODOLOGY

This section presents the architecture design of the proposed D-HAN-Net which is schematically shown in Fig. 1.

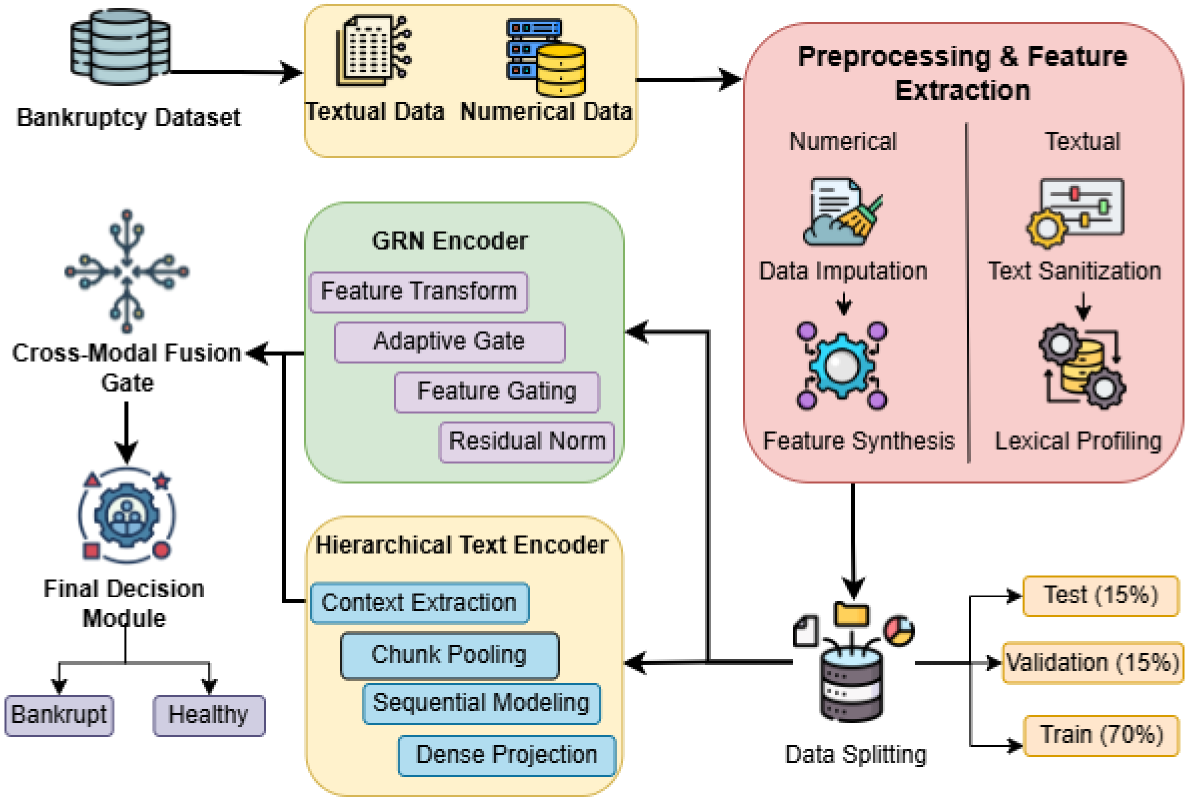

Fig. 1. End-to-end architecture of the proposed dual modality learning framework for bankruptcy prediction.

The framework consists of a combination of multimodal cohort construction, disclosure-driven risk feature engineering and a hybrid neural architecture with gated numerical encoding, hierarchical document encoding and cross-modal fusion to predict next year bankruptcy.

A. Dataset Description

This study is based on the publicly available Figshare dataset of multimodal bankruptcies [31]. In each instance there are MD&A text, audit opinion text, structured numerical attributes and the next year bankruptcy label binary (indexed by UID). The circumstances are not in our favor since bankrupts are few and financial filings are long. Therefore, the decision relevant signals are preserved by imbalance aware sampling and hierarchical text processing. Potential selection bias could also arise due to the fact that financially distressed firms have larger management reports and larger auditor opinions. Distressed companies often offer specific explanations about the reasons for liquidity situations, operational uncertainty, debt commitments, or restructuring efforts. Consequently the intensity of narrative may be partly correlated with the risk of bankruptcy. To reduce this concern, the proposed framework is built on the basis of contextual semantic representations that are derived from segmented disclosures rather than using raw document length as a direct predictor. Long disclosures are segmented into multiple chunks to reduce excessive truncation while preserving contextual semantic information across lengthy financial reports.

B. Data Preprocessing

The pipeline begins by reading both of the raw sources in chunks of size 5000 avoiding memory overflow and then aligning them by UID with an inner merge so that every retained record contains both modalities. In order to minimize the imbalance while preserving all the positive samples, all bankrupt UIDs are kept. To address the serious class imbalance typical of bankruptcy data, all bankrupt companies are kept but a sample of healthy companies is randomly selected in a ratio of around 5:1. This sampling strategy mitigates the extreme imbalance without having too few observations from the majority class for stable model training. The last cohort is divided by stratified sampling into train 70%, validation 15% and test 15%, which makes the class ratios stable between partitions and prevents the statistical instability that time-based splits would cause due to the extreme rarity of bankruptcies. The resulting distribution of healthy and bankrupt firms in the training, validation and test sets can be seen in Fig. 2.

Document segmentation is carried out for long financial disclosures. After removing noise in the form of the HTML document, and common boilerplate text, each document gets tokenized and broken into fixed sized chunks. Let T be the number of tokens in a document and L be the chunk's max token length. The number of chunks N is calculated by dividing the document into segments of length L, as in Eq. (1). In our experiments the number of tokens in a chunk is 512, following the choice of the maximum sequence length of the transformer encoder. Each document is further reduced in size to a maximum of 20 chunks, to manage the memory

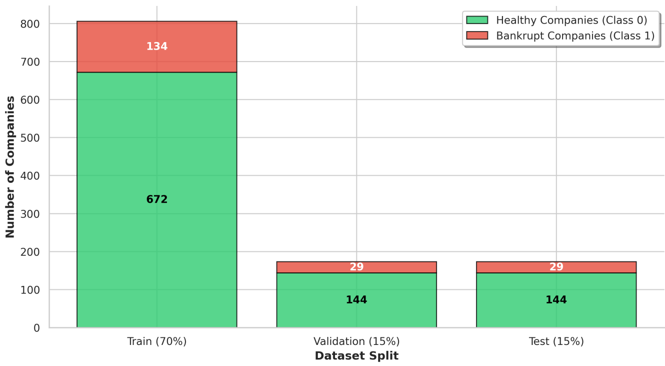

Fig. 2. Class distribution of healthy and bankrupt firms in the training, validation, and test sets.

usage while maintaining enough contextual information. While this memory-constrained limit truncates 67.01% of the longest documents and yields an overall truncation loss of 33.62%, it securely retains the most critical semantic cues and distress indicators, which typically appear in the earlier sections of management and auditor reports.

$$C = \min\left(C_{max}, \lceil \frac{T}{L} \rceil\right) \tag{1}$$

This chunked representation prevents hard truncation and supports hierarchical encoding.

C. Feature Engineering and Selection

A multivariate feature set is created to represent the accounting signals as well as the narrative risk cues. All columns other than UID and label, item 7, opinion text and the combination of raw text field are treated as candidate numerical predictors. Object type variables are label encoded, Infinite values are replaced by missing, Variables with missing rate higher than 0.60 are removed. Remaining missing entries are imputed using an iterative multivariate imputation maintaining cross feature dependencies.

To explicitly capture distress language, a unified disclosure field is formed by concatenating item 7 with audit opinion text. For a lexicon V and the token set W_i of document i, the density feature is defined in Eq. (2):

$$Density_{V,i} = \frac{\sum_{w \in W_i} I(w \in V)}{|W_i|} \tag{2}$$

Using negative, litigious, and uncertainty lexicons, we construct three density indicators, and a binary going concern flag is added when the phrase is present. These linguistic red flags strengthen interpretability and support later robustness reasoning.

Mutual information is then employed to generate the top 5 numerical predictors and subsequently only the interaction-only second-degree polynomials terms are produced. This expansion of the number of base predictors to 5 is the limit that is set so as to guarantee efficiency in the computation and to avoid overfitting. Any further extension would further raise the feature space exponentially, which is necessarily going to incur the curse of dimensionality and terrible multicollinearity, due to the extreme imbalance between classes in the case of bankruptcy data. The model only uses the top 5 predictors which is a sufficient approach to capturing the most statistically significant non-linear relationships with no unnecessary noise being modeled. For two selected predictors x_a and x_b the interaction is defined in Eq. (3):

$$x_{ab} = x_a x_b \tag{3}$$

All the numerical predictors are clipped at 1st and 99th percentile and transformed into a normal like distribution through quantile transformation. A final top K predictor is then selected by mutual information which is used by the model. The sensitivity analysis is done to find the best feature selection and test the strength of the model by changing the value of K (e.g., K = 10, 20, 30, and 40). The last model uses K=20 due to the maximum overall validation performance. The impact of this option is also addressed in the experimental results.

D. Model Architecture of the Hybrid D-HAN-Net

The proposed architecture is designed to learn stable numerical structure, preserve long range disclosure evidence, and align both modalities through attention guided fusion. The pipeline contains four stages.

1) Hierarchical numerical encoder: The GRN framework supports nonlinear transformation and performs the original feature representation by residual connections. Financial ratios tend to be heterogeneously distributed and non linearly interacting. Thus, their predictive structure cannot be well represented by a simple linear transformation. In order to stabilize the learning process, the encoder uses two nonlinear transformation layers and then dropout regularization and projection residual. A gated residual network is used to project the selected predictors to an embedding of 128 dimensions. Let r be the projected residual and $f(x)$ be a nonlinear transform, then the numerical representation is defined in Eq. (4). The hidden representation dimension is set to 128 in our experiments.

$$h_{num} = BN(r + g \odot f(x)) \tag{4}$$

where, $g = \sigma(W_g f(x))$ and $\odot$ denotes element-wise multiplication. This gating mechanism enables the network to adaptively suppress noisy financial indicators while amplifying features that strongly correlate with bankruptcy risk. Batch normalization further stabilizes training by normalizing the representation distribution.

2) Hierarchical text encoder: The FinBERT application is informed by the fact that corporate disclosure is a domain-structured linguistic. Accounting terminology, expressions of financial risk, and statements of uncertainty are common management discussion narrative language that is not well reflected in general language corpora. FinBERT is trained on financial text, and thus it picks up contextual semantics to do with financial distress, compared to general-purpose BERT models. The GRU layer is bidirectional and additional to this, the layer performs segment level embedding on the entire document and this identifies long-range dependencies among disclosure segments. The contextual embedding of each segment is obtained from the pooled output of FinBERT, which is chosen over standard BERT models to accurately interpret the specialized financial terms and unique phrasing found in management narratives. These segment embeddings are then organized as a sequence and passed through a bidirectional GRU layer. The bidirectional recurrent layer aggregates contextual information from both earlier and later segments, enabling the model to capture narrative dependencies across the processed sequence.

A bidirectional GRU aggregates chunk embedding into a single 128-dimensional representation h_{text}. The FinBERT parameters are frozen during training to reduce computational cost and prevent overfitting on the relatively small bankruptcy dataset. The resulting sequence representation is aggregated by the GRU layer to form a document-level embedding. Compared with fully transformer-based long-document architectures, the bidirectional GRU offers lower computational complexity and memory overhead. It also provides stable sequence aggregation for small and highly imbalanced financial datasets.

3) Cross-modal fusion with gating: Once numerical and textual representations are obtained, the next stage aligns both modalities through a cross-modal attention mechanism. Bankruptcy signals often emerge when textual statements confirm or contradict financial ratios. The fusion module therefore enables interaction between these two sources. The cross-attention module uses the summarized textual representation h_{text} as a Query to attend to the numerical representation h_{num}, acting as Key and Value, producing an attention-refined text vector h_{attn}. A learnable fusion gate then adaptively blends both streams, as defined in Eq. (5):

$$h_{fuse} = z \odot h_{num} + (1 - z) \odot h_{attn} \tag{5}$$

where, $z = \sigma([h_{num}; h_{attn}])$, the gate value z lies in the range 0 to 1, where values closer to 0 indicate stronger reliance on textual signals, whereas values closer to 1 indicate greater reliance on numerical indicators. Here, σ denotes the sigmoid activation function, $[;]$ denotes concatenation, and $\odot$ indicates element wise multiplication. This mechanism allows the model to adaptively decide whether bankruptcy evidence should rely more strongly on financial ratios or on textual disclosure signals for each firm. To test the effectiveness of the proposed cross-modal fusion gate, an easier to implement late fusion baseline is also considered. In this alternative environment the numerical and textual representations are simply concatenated and fed to the classifier without interaction between the two representations based on attention. This comparison is helpful for determining if the proposed fusion mechanism offers a meaningful advantage over a simpler multimodal integration strategy.

4) Probabilistic prediction and decision logic: The fused representation is finally passed through a shallow classifier consisting of a fully connected layer with GELU activation followed by dropout regularization. This stage converts the fused feature vector into a scalar bankruptcy probability. The final prediction is obtained using a sigmoid activation, defined in Eq. (6):

$$p = \frac{1}{1 + e^{-(w^\top h_{fuse} + b)}} \tag{6}$$

where, p represents the predicted probability that the firm will experience bankruptcy in the next fiscal period. The sigmoid function ensures that the output remains within the range of zero to one. During evaluation, the predicted probability is compared with an optimal decision threshold determined from the validation precision recall curve. This adaptive threshold selection helps balance precision and recall under severe class imbalance.

E. Training Strategy and Optimization

The training procedure is designed to remain stable under extreme class imbalance and long document variability, while ensuring that improvements in AUC and F1 are not driven by a single favorable split. The overall objective is to learn a calibrated bankruptcy probability that is robust across random initializations and sampling noise.

1) Imbalance-aware training and loss optimization: Bankruptcy samples are rare, hence standard mini batch sampling would lead to overfitting of the majoritarians of the healthy sample class by the optimizer. To address this, a technique of weighted random sampling is applied, at the instance level. Each training example is assigned a weight inversely proportional to its class frequency and batches are drawn with a WeightedRandomSampler. This makes the effect of bankrupt instances more effective without duplicating them deterministically, so the gradient bias for the negative class decreases. Model parameters are optimized using AdamW, which combines adaptive learning rates with decoupled weight decay.

A learning rate of 1.5×10^{-5} is selected to maintain stable updates when training with frozen language representations and a relatively small classifier head. Dropout is applied in the numerical encoder, fusion feedforward layers, and classifier head to reduce co adaptation between modalities and to prevent the fusion gate from collapsing to a single modality. Even with balanced sampling, many bankrupt signals remain subtle, so the training objective must prioritize difficult cases. The model is trained using focal loss, which down weights easy samples and focuses learning on hard or misclassified examples. With $y \in \{0, 1\}$ and predicted probability p, define $p_t = p$ if $y = 1$ else $1 - p$. The loss is defined in Eq. (7):

$$L_{focal} = -\alpha(1 - p_t)^{\gamma} \log(p_t) \tag{7}$$

The focusing parameter is set to $\gamma = 2$ in our experiments. The balancing factor $\alpha = 0.25$ is used to assign higher importance to the minority bankrupt class.

2) Validation protocol and stability analysis: Training runs are for up to 40 epochs, but performance is not assumed to improve monotonically. The validation AUC is calculated after every epoch with the help of predicted probabilities. The model checkpoint with the highest validation AUC value is saved and early stopping is triggered when there is no improvement in validation AUC for 10 consecutive epochs. This helps avoid over training the classifier head and the fusion layers which is particularly common when the text encoder is frozen and the capacity is focused on fusion block. Since the dataset is imbalanced, applying a default classification threshold of 0.5 often yields a high AUC but a poor F1-score for the positive class.

Therefore, after being trained, the decision threshold is selected by maximizing the F1 score over the precision recall curve, which results in an equilibrium operating point between false alarms and missed bankruptcies. This threshold is then applied to the held out test set for final reporting. In order to guarantee the robustness, all experiments are repeated under multiple random seeds, which alters the selections of healthy samples and initializations of fusion and classifier parameters. Final results are presented in terms of the mean and standard deviation, which indicate the stability of the model with regard to sampling uncertainty and randomness of optimization.

F. Evaluation Protocol and Robustness Reporting

The assessment plan will be used to facilitate fair comparison of the offered model and the baseline methods. Stratified sampling is used to partition the dataset into training, validation and test sets such that the classes are always distributed equally across the partitions. The selection of models is done on the basis of validation performance in which the primary monitoring measure is the Area Under the ROC Curve. Due to the high imbalance in bankruptcy prediction, it does not mean that the eventual classification point is 0.5. Rather, the maximum value of the F1 score on the validation set is chosen and used on the held out test set. This technique offers a better balance between trade off in precision and recall. In order to measure robustness, each experiment is run with a number of random seeds and the final value is presented as mean and standard deviation. Besides, the suggestion model is contrasted with the methods of strong baseline to demonstrate the efficiency of the multimodal architecture.

IV. Results and Discussion

A. Experimental Configuration and Implementation Details

All experiments have been implemented in Python. Numerical baselines were built using Scikit-learn, the proposed multimodal model was implemented using PyTorch, and transformer baselines were implemented using HuggingFace Transformers. Experiments were done on the Kaggle research platform with GPU/CUDA to speed up the training and inference. Data preprocessing and feature engineering were done using Pandas, NumPy and Scikit-learn. Performance was measured in terms of Accuracy, Precision, Recall, F1-score and AUC. To make the comparison as broad as possible, baseline models were developed for both the numerical and textual modalities. The number baselines were Random Forest 100 trees, XGBoost logistic objective, SVM probabilistic output, and a 2-layer multilayer perceptron with 64 and 32 hidden layers. The textual baselines were BERT-Base, RoBERTa, DistilBERT, and ELECTRA fine-tuned for three epochs and batch 16 maximum sequence length 128.

D-HAN-Net was optimized using AdamW with learning rate of 1.5×10^{-5}, batch size 16 and max. 40 epochs. Early stopping was used according to validation AUC using patience of 10 epochs. To deal with the class imbalance, a weighted random sampling was adopted in the mini-batch construction and focal loss was taken into consideration to give priority to difficult samples. The final classification threshold was chosen from the precision-recall curve by finding the F1-score (Table I).

TABLE I. Key Training Configuration for the Proposed D-HAN-Net Model.

Parameter	Value
Healthy:Bankrupt Ratio	5:1
Data Split (Train/Val/Test)	70% / 15% / 15%
Selected Numerical Features	20
Text Encoder Aggregation	Bidirectional GRU
Numerical Representation	Gated Residual Network (GRN)
Multimodal Fusion	Cross-modal Attention + Gate
Optimizer & Learning Rate	AdamW, 1.5×10^{-5}
Batch Size & Max Epochs	16, 40
Imbalance Handling	Weighted Sampling, Focal Loss ($\gamma = 2$)

B. Performance Evaluation and Empirical Analysis

Across all baselines, the test-set results confirm that bankruptcy signals are captured much more reliably by structured financial variables than by narrative text alone, as illustrated in Table II. Among numerical baselines, MLP achieves the best overall score with accuracy 0.9133 and AUC 0.9672 and SVM and Random Forest closely follow with accuracy close to 0.89-0.90 and AUC above 0.94. However, Table II also shows a clear trade-off in which XGBoost and SVM have the highest recall of 0.9310 but their precision remains lower with 0.5510 and 0.6136, which implies a higher number of false positive bankruptcy alarms. The transformers that are text-only perform a weaker job overall, with accuracy maximum of 0.8324 and AUC mostly in between 0.61-0.69 which means that separability with text only is limited. D-HAN-Net is unique in the way that it enhances the overall ranking and classification simultaneously. It has the highest accuracy of 0.9400 and the best AUC of 0.9709, which shows the most obvious separation between bankrupt and healthy firms across thresholds in Table II. At the same time, it maintains precision of .7485 and recall of .8828 which means the model does not face the common tradeoff between high recall and many false positives. This balance can be seen in its top F1-score in comparison with all baselines. The result is consistent with recent multimodal bankruptcy studies which indicate that combining financial and textual information improves minority distress detection [36], [37].

Table III provides a comparison between the AUC scores of the model with different sets of features (K). When only 10 features were used, the AUC is less (0.8997) since the model has limited depth with regards to finances. Maximum performance is achieved at the top of the optimal K = 20 (our proposed model) which captures the most effective and relevant financial signs of risk assessment. Further increases in AUC to 30 and 40 bring the values to 0.9373 and 0.9303, respectively. This deterioration confirms that too many variables bring about noise and redundancy of data which ultimately diminishes the total accuracy of prediction.

Fig. 3 reports the confusion matrix of D-HAN-Net on the held out test set and explains how the final accuracy

TABLE II. BASELINE COMPARISON RESULTS ON THE TEST SET

Type	Model	Accuracy	Precision	Recall	F1-Score	AUC
Numerical	Random Forest	89.60%	64.10	86.21	73.53	95.31
Numerical	XGBoost	86.13%	55.10	93.10	69.23	93.92
Numerical	SVM	89.02%	61.36	93.10	73.97	95.47
Numerical	MLP	91.33%	68.42	89.66	77.61	96.72
Text	BERT-Base	83.24%	50.00	37.93	43.14	68.51
Text	DistilBERT	67.63%	27.87	58.62	37.78	66.55
Text	ELECTRA	73.99%	30.95	44.83	36.62	61.09
Numerical +Text	D-HAN-Net	94.00%	86.00	92.00	88.00	97.34

TABLE III. AUC PERFORMANCE BASED ON FEATURE SELECTION (TOP_K).

Number of Selected Features (TOP_K)	AUC Score
TOP_K = 10	0.8997
TOP_K = 20	0.9734
TOP_K = 30	0.9373
TOP_K = 40	0.9303

gets formed using class wise outcomes. The confusion matrix shows that 135 healthy firms are correctly classified as healthy and 26 bankrupt firms are correctly identified as bankrupt. The remaining errors are limited, and asymmetric in a good way for use in early warning. Only 9 healthy firms are flagged as bankrupt which represents false alarms which may trigger unnecessary investigation, while only 3 bankrupt firms are classified as healthy which represents missed distress cases which are usually more costly in practice. The low number of false negatives shows that the model captures most of the distressed firms, and does not depend on the majority class to inflate the accuracy. At the same time, a low number of false positives means the model does not create over sensitive screening that will overwhelm stakeholders with alerts. Overall, in the matrix, D-HAN-Net is seen to have a practical balance between caution and precision that generates high true positive detection and keeps the misclassification rates under control. This behavior is consistent with a robust multimodal decision process that relies on financial ratios, as well as disclosure signals.

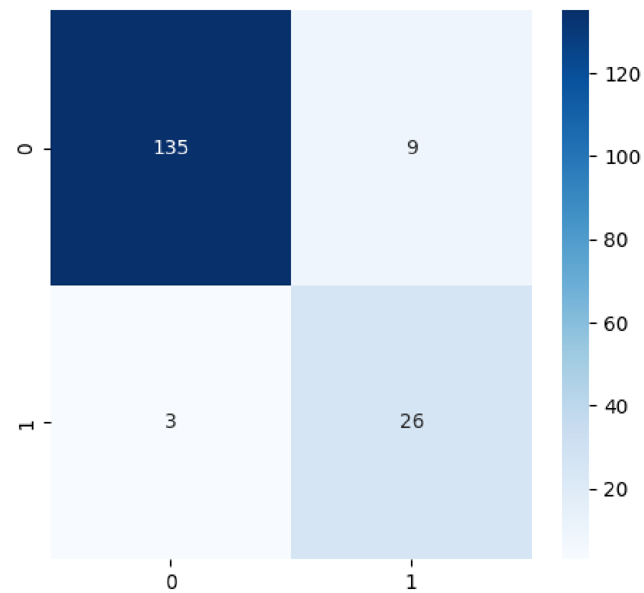

Fig. 3. Confusion matrix of the D-HAN-Net on the test dataset.

C. Threshold Independent Performance Evaluation

Instead of improving performance at a single fixed threshold, D-HAN-Net provides a predictable better performance balance across a wide range of operating points, and this can be seen from both of the curves in Fig. 4 and Fig. 5. In Fig. 4, its ROC trajectory remains most close to the top left region over most of the false positive range, which represents the best separating capability between bankrupt and non-bankrupt firms. The early rise of the curve indicates that the sensitivity is gained rapidly and the small drift to the right indicates that false positives are still under control. The numerical baselines lag behind them with very high but less high curvature so their extra true positive gain tends to demand more false positives. The text-only transformers are still closer to the diagonal trend, with very limited discriminative power across thresholds in Fig. 4. This suggests that narrative disclosures alone may capture uncertainty-related semantic cues but remain insufficient for stable bankruptcy discrimination without supporting financial indicators. D-HAN-Net is unique in that it enhances overall ranking as well as classification together. It has the highest accuracy of 0.9400 and the best AUC of 0.9734 demonstrating the clearest separation of bankrupt and healthy firms for different thresholds, as in Table II. At the same time, it maintains the precision level at 0.86 and recall level at 0.88 and thus the model does not suffer from the usual tradeoff where higher recall comes with a lot of false positives. This balance is reflected in its top F1-score compared with all the baselines. The improved performance likely arises from the complementary interaction between structured financial indicators and contextual narrative evidence under highly imbalanced distress conditions.

The same advantage passes on to the more decision critical view in Fig. 5. D-HAN-Net maintains precision when recall is increased, thus can recall more bankrupt firms without quickly raising false alarms. This is important in screening tasks where it is costly to miss a bankrupt case, but too many false alerts will also make the tool less practical. Several numerical baselines exhibit a steeper drop in precision at higher recall, indicating that the sensitivity (i.e. pushing to create more alarms) produces more noisy alarms. The text-only models fall to low precision early, and are weak throughout the curve, indicating that the narrative signals alone are insufficient to

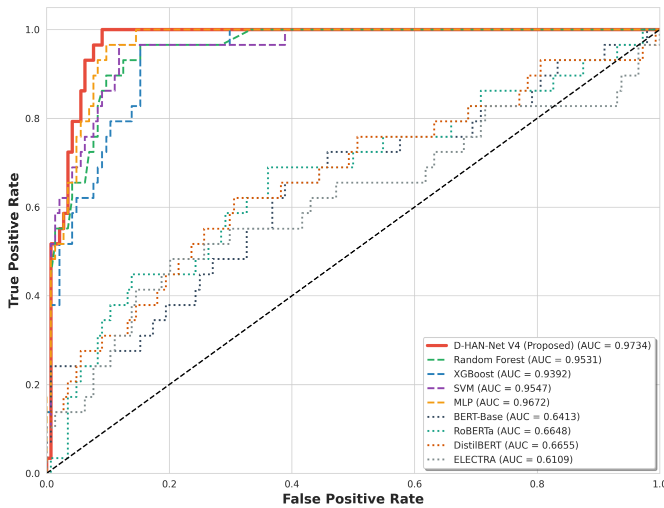

Fig. 4. Receiver Operating Characteristic (ROC) curves for all models.

secure stable retrieval.

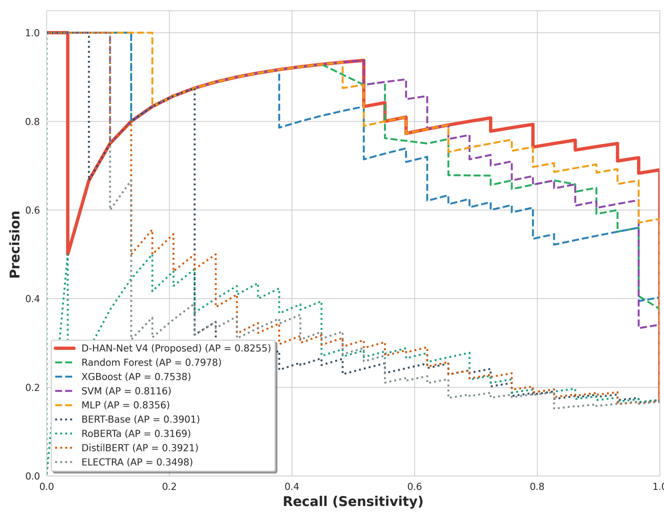

Fig. 5. Precision–Recall (PR) curves for all models.

Taken together, Fig. 4 and Fig. 5 imply that D-HAN-Net has strong ranking quality and stable class retrieval, simultaneously, and the advantage is retained in the whole threshold space instead of only one chosen decision point, which is helpful for deployment under various risk preferences.

D. Model Robustness and Cross Seed Stability Analysis

Robustness analysis evaluates whether D-HAN-Net stays reliable when training is repeated with different random seeds and compares this stability against strong numerical baselines in Table IV. The goal is to ensure that the reported gains are not caused by one favorable run, but remain consistent across multiple executions of the pipeline. Table IV indicates that D-HAN-Net achieves the best overall performance across five seeds. It records the highest accuracy at 94.00 ± 0.57 and the strongest AUC at 97.34 ± 0.35, showing superior separation between bankrupt and healthy firms. The F1 score is also the highest at 88.00 ± 0.75, supported by high precision at 86.00 ± 4.14 and high recall at 92.00 ± 4.68, which reflects balanced detection rather than a one sided gain. The error profile further strengthens this conclusion in Table IV. D-HAN-Net shows the lowest Type I error at 6.11 ± 1.55, which means fewer false bankruptcy alarms than XGBoost, Random Forest, and SVM. Type II error remains controlled at 11.72 ± 4.68, while Random Forest shows a higher Type II error at 17.93 ± 4.02, implying more missed bankrupt cases. XGBoost and SVM reach very high recall at 93.10, but their lower precision and higher Type I error indicate a larger false positive burden. Overall, Table IV supports that combining auditor text with financial ratios produces a more stable and dependable predictor than single modality baselines.

E. Component Contribution and Ablation Investigation

The ablation study is to break the full D-HAN-Net into key parts, to quantify the contribution of each block to the prediction quality. This type of analysis helps validate the components, as well as measuring the synergy, that the overall gains are the result of the combined design, rather than one block alone. As presented in Table V, the result of the full model is the best, which is accuracy 94.00, F1-score is 88.00 and AUC is 97.34. When GRN is removed, performance deteriorates drastically to accuracy 83.82 and AUC 84.77, suggesting that gated refinement is central for strong representation learning. Removing sequential aggregation also decreases performance, with accuracy 84.97 and F1 64.79, which suggests that sequence level cues help the model form a more stable decision. When attention fusion is removed, F1 drops to 61.76, AUC drops to 86.69, this means that cross-modal alignment is required to maintain a balance between the predictions. The biggest degradation occurs when focal loss is removed, in which F1 will be 60.38 and AUC will be 83.33, indicating the importance of the loss design in handling class imbalance and avoiding unstable learning. Overall, Table V suggests that strong and consistent results demand the complete integration of GRN, sequential modeling, attention fusion and focal loss, as opposed to some simplified variant.

F. Model Interpretability Through Feature Impact Analysis

A clear hierarchy of risk drivers can be seen from the SHAP summary in Fig. 6. Going Concern Opinion is the best contributor, in which higher feature values drive the predictions into the bankrupt category and lower values are more likely to be healthy companies. The extensive dispersion of SHAP values indicates that audit opinion is a powerful separating signal particularly to firms whose explicit distress is enduring. The second and third in pecking order are retained earnings and variables that are associated with net income, which implies that reduced profitability and exhausted accumulated earnings add to the risk of bankruptcy. Most of their SHAP effects are skewed towards the positive in unfavorable values indicating that the model considers the weaknesses of earnings as a direct distress signal. It is also indicated by Fig. 6 that the model is based on the mix of audit signals and core indicators of financial strength instead of the particular ratio. Balance sheet variables like current assets and total stockholders equity have a similar direction with less financial

TABLE IV. Robustness Analysis Across 5 Random Seeds

Model	Accuracy	AUC	F1-Score	Precision	Recall	Type I Error
XGBoost	86.13 ± 0.00	93.92 ± 0.00	70.89 ± 0.00	55.10 ± 0.00	93.10 ± 0.00	15.28 ± 0.00
Random Forest	90.06 ± 0.77	94.95 ± 0.36	74.93 ± 0.75	66.78 ± 3.33	82.07 ± 4.02	8.33 ± 1.58
SVM (RBF)	89.02 ± 0.00	95.57 ± 0.11	75.68 ± 0.00	61.36 ± 0.00	93.10 ± 0.00	11.81 ± 0.00
D-HAN-Net (Proposed)	94.00 ± 0.57	97.34 ± 0.35	88.00 ± 0.75	86.00 ± 4.14	92.00 ± 4.68	6.11 ± 1.55

TABLE V. Ablation Study of D-HAN-Net: Impact of Key Components on Performance.

Model Variant	Accuracy	F1-Score	AUC
D-HAN-Net	94.00%	88.00	97.34
w/o GRN	83.82%	61.97	84.77
w/o Sequential Agg (Bi-GRU)	84.97%	64.79	89.22
w/o Attention Fusion	84.39%	61.76	86.69
w/o Focal Loss	87.28%	60.38	83.33

strength being associated with increased risk. Other accounting indicators such as federal tax reconciliation and comprehensive income also have effects on prediction, but the average effects are smaller. This indicates that the model reflects primary distress indicators and secondary accounting indicators linked to declining performance. Altogether, the feature effects are economically sound. The model relates the indicators of audit-based concerns with the profitability and solvency indicators within one interpretable decision framework, with the highest impact on the audit opinion, then the quality of earnings, and the support provided by the resilience of the balance sheet, respectively, to the ultimate prediction.

G. *Case-Specific Evaluation and Linguistic Interpretability*

This case study highlights how confidently each of the models flags a difficult bankrupt firm in Table VI. D-HAN-Net has the highest bankrupt confidence of 61.67, showing the clearest risk judgment of this sample among all methods. SVM is close at 60.93, while XGBoost is lower at 55.18, which seems to exhibit lower level of commitment despite risk detection. A larger gap is seen for the rest of the baselines in Table VI. Random Forest drops to 44.00 and MLP drops even further to 32.07, meaning that they consider the same firm less decisively bankrupt. The lowest scores are experienced by the text only transformers, hovering around 13.12 to 16.38, suggesting that they mostly fail to pick up on the evidence of distress in this difficult example. Overall, Table VI presents that D-HAN-Net is the most decisive model on a hard case, where single modality baselines tend to under estimate the bankruptcy risk.

Fig. 7 illustrates the most common linguistic cues to financially struggling firms taken from Management Discussion and Analysis and auditor opinion sections. The word cloud emphasises such dominant words as costs, credit, interest, sales, capital, and accounts that are often found in bankruptcy related disclosures. The prominence of these terms reflect common financial distress narratives including rising operational costs, debt obligations, declining revenues, and restructuring pressures. This visualization is qualitative evidence that there are meaningful warning signals of corporate failure in the textual disclosures. Such linguistic patterns complement the numerical financial indicators that are to be used in the proposed framework and help to explain the role of textual evidence in predicting bankruptcy. The analysis supports the interpretability objective it shows that the model captures real world distress related language present in corporate reports rather than relying solely on statistical correlations.

Fig. 6. SHAP Summary plot: Evaluating global feature importance and directional impact in D-HAN-Net.

TABLE VI. Critical Case Study: Cross-Model Bankruptcy Confidence Scores (Test Index = 45).

Model Architecture	Confidence Score (Bankrupt)
SVM	60.93
XGBoost	55.18
Random Forest	44.00
MLP	32.07
BERT-Base	15.49
ELECTRA	16.38
DistilBERT	13.12
D-HAN-Net (Proposed)	61.67

TABLE VII. Benchmark Comparison of Existing Models with the Proposed D-HAN-Net

First Author	Year	Architecture	Dataset	AUC
Ainan et al. [34]	2024	Hybrid Machine Learning	Unbalanced Polish Dataset	0.9580
Mancisidor et al. [37]	2022	CMMD	U.S. Firms	0.9327
Fasano et al. [32]	2024	Explainable AI	Corporate Financial Dataset	0.9270
Arno et al. [36]	2023	Multimodal Ensemble	ECL Dataset	0.9150
Correia [35]	2025	Traditional Accounting Models	CRSP U.S. Firms	0.9136
Xiao et al. [33]	2024	DGCN-TA	Enterprise Relationship Networks	0.8867
Mai et al. [7]	2019	Deep Learning	U.S. Public Companies	0.8560
D-HAN-Net (Proposed)	2026	D-HAN-Net	Multimodal	0.9734

Fig. 7. Linguistic red flags extracted from bankrupt firms' auditor reports.

Table VII indicates that the proposed D-HAN-Net is an effective tool in comparison to some other recent studies on bankruptcy prediction. Prior deep learning studies like Mai et al. had already achieved an AUC of 0.8560, which points to a moderate level of predictive performance in the case of using textual disclosures as the main predictive factor. Subsequent methods were able to enhance performance by adding more structural or financial data. As an illustration, the DGCN-TA model by Xiao et al. resulted in an AUC of 0.8867 with the classical accounting based model by Correia et al. having an AUC of 0.9136. Further developed architectures also improved the performance of prediction. Arno et al. multimodal ensemble method achieved an AUC of 0.9150, and Fasano et al. showed the applicability of explainable AI of 0.9270. On the same note, the CMMD model of Mancisidor et al. scored 0.9327, and the hybrid machine learning model of Ainan et al. scored 0.9580, which is among the highest scores ever seen among previous researchers. Although these gains have been made, the proposed D-HAN-Net has the best AUC of 0.9734, which is obviously above all current methods, as summarized in Table VII. This outcome shows that the hierarchical multimodal structure better extracts intricate bankruptcy indicators from textual and financial features. By combining these complementary modalities, the proposed framework achieves superior discrimination and more reliable predictive performance. This strength is evidenced by the model's robustness and consistently strong predictive capability across numerous experiments.

V. Conclusion

A robust early warning system for corporate financial distress is essential for investors, regulators, and financial institutions to manage risk and preserve market stability. This study developed D-HAN-Net, a hybrid dual-stream deep learning architecture that integrates structured financial indicators with textual signals derived from corporate disclosures. By jointly modeling numerical financial dynamics and narrative evidence from MD&A sections and auditor reports, the proposed framework captures complementary distress signals that are often overlooked in single-modality approaches.

The empirical results show that the combination of financial ratios and textual information from written disclosure is able to improve the bankruptcy prediction performance significantly. The proposed model has the highest performance in terms of all evaluation metrics such as accuracy, AUC, and F1-score with an efficient balance between precision and recall. Compared with the standard machine learning benchmarks and the transformer-based text models, D-HAN-Net is able to reduce classification errors and increase the overall classification reliability. Moreover, the fact that results are consistent across multiple random seeds suggests that the observed performance gains are robust, rather than sample-specific. The ablation analysis further indicates that every architectural component has a material contribution to the overall predictive performance of the model, with especially high contribution of the gated residual numerical encoder, hierarchical text representation, and cross-modal fusion mechanism. These results highlight the importance of bringing diverse financial information together in a cohesive predictive model. In addition, the interpretability analysis indicates that the model captures economically meaningful indicators of distress. Audit-related going concern indicators as well as profitability and balance sheet strength indexes become key variables. The linguistic analysis of disclosure texts also suggests that the model finds distress-related narrative patterns often found with financially vulnerable firms.

Beyond the predictive performance, the results of this study have meaningful implications for the financial reporting regulation. Narrative disclosures, especially MD&A discussions and auditor reports, provide informative signals of financial distress. Therefore, regulatory authorities may wish to consider enhancing mandatory reporting standards to increase the clarity, consistency and decision-usefulness of such disclosures. More standardized and informative reporting practices could help improve the effectiveness of data-driven early warning systems and provide support to more timely oversight by regulators, creditors and investors.

Overall, the combination of structured financial data and narrative disclosure is a more complete representation of corporate financial condition. The proposed architecture provides a workable and reliable framework for early prediction of bankruptcy, as well as informed financial risk assessment. Despite the strong predictive performance, the proposed framework may still be affected by temporal distribution shifts, industry-specific reporting variations, and incomplete disclosure availability in real-world financial environments. Future research may investigate temporal forecasting settings and robustness under changing economic conditions.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Southeast University, Dhaka, Bangladesh, for providing the essential laboratory facilities and technical support required to conduct this research.

REFERENCES

[1] Y. Ji, X. Zhou, Z. Cai, and Z. Yin, "Enhancing status-based optimization via dual dispersal mechanisms for feature selection in bankruptcy prediction," *Neurocomputing*, vol. 668, p. 132394, 2026. doi: 10.1016/j.neucom.2025.132394.

[2] Y.-C. Lin, R. Padliansyah, Y.-H. Lu, and W.-R. Liu, "Bankruptcy prediction: Integration of convolutional neural networks and explainable artificial intelligence techniques," *International Journal of Accounting Information Systems*, vol. 56, p. 100744, 2025. doi: 10.1016/j.accinf.2025.100744.

[3] W. Liu, H. Wu, T. Wang, et al., "Elite elimination osprey optimization algorithm optimized kernel extreme learning machine for bankruptcy prediction problems," *Scientific Reports*, vol. 16, p. 6246, 2026. doi: 10.1038/s41598-026-37249-9.

[4] K. Jiang, X. Zhao, Y. Li, et al., "A thinking innovation strategy based Northern goshawk optimizer enhanced extreme learning machine for bankruptcy prediction problems," *Scientific Reports*, vol. 16, p. 9628, 2026. doi: 10.1038/s41598-025-34452-y.

[5] G. Lombardo, M. Pellegrino, G. Adosoglou, S. Cagnoni, P. M. Pardalos, and A. Poggi, "Machine learning for bankruptcy prediction in the American stock market: Dataset and benchmarks," *Future Internet*, vol. 14, no. 8, p. 244, 2022. doi: 10.3390/fi14080244.

[6] G. Ranjbaran, D. R. Recupero, G. Lombardo, and S. Consoli, "Leveraging augmentation techniques for tasks with unbalancedness within the financial domain: a two-level ensemble approach," *EPJ Data Science*, vol. 12, no. 1, p. 24, 2023. doi: 10.1140/epjds/s13688-023-00402-9.

[7] F. Mai, S. Tian, C. Lee, and L. Ma, "Deep learning models for bankruptcy prediction using textual disclosures," *European Journal of Operational Research*, vol. 274, no. 2, pp. 743–758, 2019. doi: 10.1016/j.ejor.2018.10.024.

[8] W. J. Mayew, M. Sethuraman, and M. Venkatachalam, "MD&A disclosure and the firm's ability to continue as a going concern," *The Accounting Review*, vol. 90, no. 4, pp. 1621–1651, 2015. doi: 10.2308/accr-50983.

[9] F. Li, "Annual report readability, current earnings, and earnings persistence," *Journal of Accounting and Economics*, vol. 45, no. 2-3, pp. 221–247, 2008. doi: 10.1016/j.jacceco.2008.02.003.

[10] T. Loughran and B. McDonald, "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," *The Journal of Finance*, vol. 66, no. 1, pp. 35–65, 2011. doi: 10.1111/j.1540-6261.2010.01625.x.

[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.

[12] T.-K. Chen, H.-H. Liao, G.-D. Chen, W.-H. Kang, and Y.-C. Lin, "Bankruptcy prediction using machine learning models with the text-based communicative value of annual reports," *Expert Systems with Applications*, vol. 233, p. 120714, 2023. doi: 10.1016/j.eswa.2023.120714.

[13] E. I. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *The Journal of Finance*, vol. 23, no. 4, p. 589, 1968. doi: 10.2307/2978933.

[14] S. P. Kothari, X. Li, and J. E. Short, "The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis," *The Accounting Review*, vol. 84, no. 5, pp. 1639–1670, 2009. doi: 10.2308/accr.2009.84.5.1639.

[15] F. Li, "The information content of forward-looking statements in corporate filings—A naïve Bayesian machine learning approach," *Journal of Accounting Research*, vol. 48, no. 5, pp. 1049–1102, 2010. doi: 10.1111/j.1475-679X.2010.00382.x.

[16] H. Kim, H. Cho, and D. Ryu, "Corporate bankruptcy prediction using machine learning methodologies with a focus on sequential data," *Computational Economics*, vol. 59, no. 3, pp. 1231–1249, 2022. doi: 10.1007/s10614-021-10126-5.

[17] T. Hosaka, "Bankruptcy prediction using imaged financial ratios and convolutional neural networks," *Expert Systems with Applications*, vol. 117, pp. 287–299, 2019. doi: 10.1016/j.eswa.2018.09.039.

[18] M. Pellegrino, G. Lombardo, G. Adosoglou, S. Cagnoni, P. M. Pardalos, and A. Poggi, "A multi-head LSTM architecture for bankruptcy prediction with time series accounting data," *Future Internet*, vol. 16, no. 3, p. 79, 2024. doi: 10.3390/fi16030079.

[19] A. B. Hussain, S. Jingchun, B. Hussain, et al., "Cross-market bankruptcy prediction: An interpretable ensemble learning framework using SHAP analysis," *Computational Economics*, 2026. doi: 10.1007/s10614-025-11266-8.

[20] D. Liang, C.-C. Lu, C.-F. Tsai, and G.-A. Shih, "Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study," *European Journal of Operational Research*, vol. 252, no. 2, pp. 561–572, 2016. doi: 10.1016/j.ejor.2016.01.012.

[21] D.-N. Wang, L. Li, and D. Zhao, "Corporate finance risk prediction based on LightGBM," *Information Sciences*, vol. 602, pp. 259–268, 2022. doi: 10.1016/j.ins.2022.04.058.

[22] S. Carta, A. Ferreira, D. R. Recupero, M. Saia, and R. Saia, "A combined entropy-based approach for a proactive credit scoring," *Engineering Applications of Artificial Intelligence*, vol. 87, p. 103292, 2020. doi: 10.1016/j.engappai.2019.103292.

[23] S. Carta, G. Fenu, A. Ferreira, D. R. Recupero, and R. Saia, "A two-step feature space transforming method to improve credit scoring performance," in *A Two-Step Feature Space Transforming Method to Improve Credit Scoring Performance*, Springer, 2020, pp. 134–157. doi: 10.1007/978-3-030-66196-0_7.

[24] D. Liang, C.-F. Tsai, H.-Y. Lu, and L.-S. Chang, "Combining corporate governance indicators with stacking ensembles for financial distress prediction," *Journal of Business Research*, vol. 120, pp. 137–146, 2020. doi: 10.1016/j.jbusres.2020.07.052.

[25] W.-C. Lin, Y.-H. Lu, and C.-F. Tsai, "Feature selection in single and ensemble learning-based bankruptcy prediction models," *Expert Systems*, vol. 36, no. 1, 2019. doi: 10.1111/exsy.12335.

[26] Y. Qu, P. Quan, M. Lei, and Y. Shi, "Review of bankruptcy prediction using machine learning and deep learning techniques," *Procedia Computer Science*, vol. 162, pp. 895–899, 2019. doi: 10.1016/j.procs.2019.12.065.

[27] A. Mashrur, W. Luo, N. A. Zaidi, and A. Robles-Kelly, "Machine learning for financial risk management: A survey," *IEEE Access*, vol. 8, pp. 203203–203223, 2020. doi: 10.1109/ACCESS.2020.3036322.

[28] G. Lombardo, G. Trimigno, M. Pellegrino, and S. Cagnoni, "Language models fine-tuning for automatic format reconstruction of SEC financial filings," *IEEE Access*, vol. 12, pp. 31249–31261, 2024. doi: 10.1109/ACCESS.2024.3370444.

[29] M. Papík and L. Papíková, "The possibilities of using AutoML in bankruptcy prediction: Case of Slovakia," *Technological Forecasting and Social Change*, vol. 215, p. 124098, 2025.

[30] M. Kubat and S. Matwin, "Addressing the curse of imbalanced training sets: one-sided selection," in *Proceedings of the 14th International Conference on Machine Learning*, 1997, pp. 179–186.

[31] A. Sideras, K. Bougiatiotis, E. Zavitsanos, G. Paliouras, and G. Vouros, "A multimodal dataset of financial disclosures, MD&A, and audit opinions with next-year bankruptcy labels," *figshare*, Dataset, 2025. doi: 10.6084/m9.figshare.30305341.v1.

[32] F. Fasano, C. Adornetto, I. Zahid, M. La Rocca, L. Montaleone, G. Greco, and A. Cariola, "The dilemma of accuracy in bankruptcy prediction: a new approach using explainable AI techniques to predict corporate crises," *European Journal of Innovation Management*, vol. 28, no. 11, pp. 1–22, 2025. doi: 10.1108/EJIM-06-2024-0633.

[33] Y. Xiao and X. Liu, "Dynamic graph convolutional network with temporal aggregation for corporate bankruptcy prediction," *IEEE Access*, vol. 14, pp. 14985–14995, 2026. doi: 10.1109/ACCESS.2026.3656673.

[34] U. H. Ainan, L. Y. Por, Y.-L. Chen, J. Yang, and C. S. Ku, "Advancing bankruptcy forecasting with hybrid machine learning techniques: Insights from an unbalanced Polish dataset," *IEEE Access*, vol. 12, pp. 9369–9381, 2024. doi: 10.1109/ACCESS.2024.3354173.

[35] M. Correia, "Accounting and corporate failure: the evolving role of accounting information in bankruptcy prediction," *Accounting and Business Research*, vol. 55, no. 5, pp. 510–537, 2025. doi: 10.1080/00014788.2025.2516273.

[36] H. Arno, K. Mulier, J. Baeck, and T. Demeester, "From numbers to words: Multi-modal bankruptcy prediction using the ECL dataset," in *Proceedings of the Sixth Workshop on Financial Technology and Natural Language Processing*, 2023, pp. 11–21. doi: 10.18653/v1/2023.finnlp-2.2.

[37] R. A. Mancisidor and K. Aas, "Multimodal generative models for bankruptcy prediction using textual data," *arXiv preprint arXiv:2211.08405*, 2022.