Temporal-Cross-Modal Intelligence for Detecting Fraudulent Crowdfunding Campaigns

Lakshmi B S[1], Rekha K S[2]
Department of Computer Science,The National Institute of Engineering-Mysuru-Affiliated to Visvesvaraya Technological University, Belagavi, India[1, 2]
Department of Computer Science, Vidyavardhaka College of Engineering-Mysuru-Affiliated to Visvesvaraya Technological University, Belagavi, India[1]
Department of Computer Science, JSS Science and Technology, Mysuru, India[2]

***Abstract*—Reward-based crowdfunding platform fraud has now become a multimodal and temporally dynamic threat, with conventional text-only or snapshot-based detection methods ineffective at detecting more complex deceptive campaigns. In this study, a Temporal Dynamics Aware Multi-Model Fraud Detection Framework (TDMM-FDF) that simultaneously models linguistic indicators, visual discrepancies, and time behavioral changes is proposed. The framework introduces three key innovations: 1) HM4, a Hidden Method-of-Moments Markov model for modeling long-range latent transitions across campaign updates, 2) Polynomial Expansion Canonical Correlation Analysis (PECCA) for quantifying nonlinear semantic discrepancies between textual narratives and associated images, and 3) Frequency-Gated GRU (FG-GRU) which separates recurrent activations into low frequency (trend) and high frequency (anomaly) components in order to achieve higher sensitivity to abrupt fraudulent behaviors. Massive simulations on an actual Kickstarter data set prove that the given architecture outperforms classical machine learning models, sequence encoders, and transformer baselines significantly [96.4% accuracy and good calibration (ECE = 0.06) and high ROC-AUC]. The supplementary role of all modules is confirmed in ablation studies, and their qualitative analyses provide precise semantic-visual discrepancies and semantic time anomalies of fraudulent campaigns.**

Keywords*—Crowdfunding fraud detection; multimodal learning; temporal behavior modeling; cross-modal consistency analysis; blockchain-based verification

I. Introduction

Over the past years, online crowdfunding sites have transformed the world of entrepreneurship and innovation funding through a system that helps individuals and startups to raise funds through the help of a large group of supporters and cuts off the middlemen. In fact, the reward-based crowdfunding model has gained popularity; specifically, the Kickstarter platform has already assisted in the launch of hundreds of thousands of campaigns, with the billions of dollars pledged [1], [2]. This liberalization of capital has opened new possibilities to both project creators and those who support the projects. However, it has also brought great new risks.

Fraud is one of the gravest threats to crowdfunding ecosystems: creators who lie about themselves, who cannot deliver on their promises, or otherwise abuse the trust of backers. The trend of misleading campaigns based on the use of linguistic indicators, insignificant responsibility of creators, and the absence of effective control over platforms is increasing. Crowdfunding fraud is a betrayal of trust, a risk to the reputation of the platform, and compromises the future sustainability of the crowdfunding business model [3]. Nevertheless, even being significant, the analysis of fraud in this field of crowdfunding is at a comparatively young age, particularly in comparison with fraud prevention in the banking or credit-card sectors.

One of the difficulties of identifying crowdfunding fraud is time-related factors. Although much of the literature is devoted to the fixed characteristics of campaigns (e.g., funding goal, number of backers, presence of video), the movement of a campaign, how updates are displayed, how the activity of backers varies with time, and how communications by creators vary with time can provide dense information about legitimacy or dishonesty. To give an example, Bernardino et al. in their exploratory analyses of crowdfunding dynamics observed that backer and updates temporal dynamics may differentiate between successful and non-successful campaigns [4]. Nevertheless, these time indicators are still mostly unexploited in the field of fraud detection studies. According to one of the recent systematic reviews [5], the detection of fraud within the context of crowdfunding tends to ignore the time and multi-modality of behavior, concentrating on static snapshot characteristics. Multi-modal information is also another important dimension. Contemporary crowdfunding efforts usually contain text (project description, updates, and comments), images and video (media of the product or prototype), and time (frequency of updates, funding pattern, reactions of backers). Fraudsters can exploit a single modality and conceal inconsistencies in another (or both) or may conceal discrepancies between modalities or in the timing of interactions. Such cross-modal incongruity as a shiny image of the product, but slow updates or missing comments by true backers can be powerful red flags, but there is little incorporation of multiple modalities and time.

This study introduces a temporal dynamics-aware multi-modal fraud detection framework, which is explicitly implemented in a reward-based crowdfunding platform. It can be driven by the fact that there is a growing sophistication in terms of deceiving potential followers using multimedia content and time behavior by fraudulent campaigns. To overcome these difficulties, we design our framework, incorporating time modelling, multi-modal feature fusion, and cross-modal

consistency validation to have a holistic mechanism of fraud detection. Our system is based on the Hidden Method-of-Moments Markov Model (HM4), the first great component that represents sequential dependencies in campaign updates and interactions between supporters. HM4 is useful in learning such latent behavioral patterns, which differentiate legitimate campaigns and those with suspicious or erratic updating behavior patterns. This time sensitivity enables the model to sense any abnormality in the frequency of postings, development of contents and the support of posts over time.

This study was inspired by the fact that the nature of deception on crowdfunding sites has been evolving at a high rate since the phishers on the crowdfunding sites have outgrown the textual lies into the so-called temporal-multimodal deception. Although the current body of literature has devoted much attention to the so-called snapshot-characteristics these approaches are getting more and more oblivious to the high-tech rhythm of the modern fraud. A dishonest designer might be able to provide a professional impression, but have a suspicious update rate or textual content that does not match visual prototype clues. It is in dire need of a framework that not only views multiple data types but also comprehends their correlation and time dynamics. The optimization of verifying temporal behaviors should be coupled with cross-distribution checks, whereby platforms can transition to the state of proactive fraud resistance and high precision.

For fusing heterogeneous data modalities, we propose a Frequency-Gated Gated Recurrent Unit (FG-GRU) architecture. This new form of GRU inserts multi-scale frequency gating to primarily highlight meaningful temporal variations. The FG-GRU takes in three streams of inputs: 1) textual features, which have been extracted based on a Hierarchical Pattern Distillation-based approach on BERT representations to identify semantic inconsistencies, 2) visual features, which have been extracted based on ICI-based CLAHE and HOG/SIFT to take in image manipulation or reuse, and 3) temporal states by the HM4 model. The combination of those modalities allows the classifier to make resilient campaign honesty forecasts. We also use Polynomial Expansion Canonical Correlation Analysis (PECCA) in cross-modal consistency checking between textual and visual modalities. This will guarantee the semantics adherence, where the potential discrepancies will be indicated, including the presence of overly professional imagery and the incoherent or misleading textual explanations. This will make PECCA more interpretable and will increase the trust in the decision-making process of the model.

The rest of the study is structured in the following way: Section II is the review of the related literature about crowdfunding fraud, temporal modelling, and multi-modal detection. Section III will provide our methodology, such as data preprocessing, feature extraction, model configuration, and training strategy. Section IV presents our results, comparison with baselines, ablation studies and qualitative case studies. Lastly, Section V concludes and identifies the future research directions.

II. Related Work

Crowdfunding has grown into a mainstream tool of financing in its early stages, but the transparency that empowers creativity can also prove flaws on the platforms in its ease of cheating. Initial studies of crowdfunding-related fraud focused their analysis more on fixed characteristics of campaigns, such as creator profile features, language indicators, and top-level engagement mechanisms, and used classical classifiers. Lee et al. have built a labelled Kickstarter corpus and have shown that forward stepwise logistic regression using engineered campaign/creator/content features can distinguish between fraudulent and legitimate campaigns with an accuracy of 87.3%; noteworthy, they also observed the vulnerability of snapshot features that disregard the dynamics of behaviors over time [1]. As previously explained, Perez et al. expanded the field to include platforms, modalities, and reported good AUC when using text-image features and traditional ML baselines, but not with temporal signals in model selection [3]. Qu and Hou wrote a corresponding thread about textual self-contradiction in a campaign and suggested a dual BERT-mT5 pipeline with sentence designs and sentiment classification to achieve 85.26% accuracy-text only, but ignoring visual/temporal cues [6]. A PRISMA bibliometric study has enhanced the fragmentation in the field and identified precisely such gaps as a deficient temporal modelling and a deficient multi-modal integration as the key obstacles to effective fraud detection in the sites of a crowdfunding business [7].

Temporal signals are significant candidates of project success and, by implication, plausibly informative of fraud detection in the event of the patterns being out of place. In each of the 2852 projects, the analysis by Solodoha revealed non-linear effects of frequency of updates: both neglect and over-updating imply different implications on the results, hence the importance of fixed snapshots being blind to important context [8]. Though neighboring entrepreneurship literature explores uncertainty, hype, and incompleteness of decisions in crowdfunding and similar environments, it usually does not go beyond this; however, these articles encourage behavioral and time-sensitive phenomena (e.g., cadence of updates, latency between commitments and posts, dynamics of backer counts) that a fraud detector should learn [9]. Together with the preceding stream, this stream justifies the requirement of sequence-sensitive models as opposed to fixed classifiers.

Because most of these deceptive campaigns appear through professional-looking media so as to conceal suspicious text, cross-modal reasoning is paramount. Lin introduced a text-image fusion image-to-text misinformation model, which is an improvement over unimodal baselines and demonstrates that joint representations can detect subtle inconsistencies that individual modalities would be blind to [10]. This concept is refined in a few deep multimodal fake-news experiments that have contrastive objectives and optimal transportation to align and compare modalities (Shen et al.), or that have contrastive learning based on data-augmentation to harden models to distribution changes (Hua et al.) [11], [12]. Segura-Bedmar and Martinez showed that CNN-based fusion achieved competitive accuracy in Fakeddit, which proves the worth of acquired cross-modal features in comparison with concatenation alone [9]. In this area (Nasser et al.; Shen et al.), architectural forms, including late versus early fusion, attention-based alignment, and consistency scoring, are synthesized and can be directly translated to crowdfunding, although benchmarks may vary

[13], [14]. The literature confirms that explicit text-image correspondence plays a critical role in determining that there is a semantic discrepancy in digital content. Mohan et al. suggested a synergistic detection model based on TextGCN, Vision Transformers to learn intricate inter-modal association [15]. Hangloo and Arora proposed a feature fusion model that is specifically created to identify multimodal fake news through matching between divergent streams of data [16]. Huang et al. revealed that the diffusion model could be used to reveal text-image inconsistencies through visual data reconstruction based on textual prompts [17]. In addition, Kumari and Singh proposed a multimodal deep learning framework that combines different techniques of feature extraction to enhance the accuracy of the classification [18]. Fraud cannot easily take a single and fixed signature, and much more frequently is manifested in local anomalies in more or less plausible regions of behavior. The Local Outlier Factor (LOF) has been appealing to high-throughput systems since it models density-based deviations without any heavy parametric restrictions. Adesh et al. explain LOF and improved-LOF in the context of HPCC environments and focus on additional considerations of scalability and practical deployment that are relevant in the context of real-time platform defense [19].

Hybrid pipelines are being more often applied in further financial and enterprise terms to stabilize predictions. Cherif et al. provided a systematic review of credit card fraud detection with disruptive technologies and stressed on the transition to integrated deep learning models [20]. Khalid et al. suggested an ensemble machine learning methodology which involves using a combination of several classifiers to improve the screening of fraudulent transactions [21]. Ismail and Haq showed how enterprise financial fraud detection can be enhanced by addressing the features of engineering and hybrid model architecture to process unstructured information [22]. These experiments validate the application of unsupervised locality scores, e.g., LOF, to scale-down pronounced supervised classification confidence.

In modern fraud detection, signals occur in a sequence, or updates, comments, promises, but models must be able to retain long-range and short-range dependencies. Other than LSTM/GRU surveys, there are two strands that are of particular interest. First, GRU, based on architecture or training innovations offer more robust sequence encoders at restricted data. To learn spatial-temporal dependencies, Liu et al. use GRUs in a graph neural network to form GR-GNN and obtain reduced error in time-series prediction and how graphical inductive bias can stabilize recurrent learning [23]. Another article by Liu et al. is devoted to evolutionary optimization of GRU hyperparameters, where the authors report steady improvement as compared to vanilla GRUs in sequential prediction [24]. Second, there are spatial-temporal models of GAST (graph attention + temporal forecasting) which show that the attention toward changing relational structure enhances predictive fidelity when encountering the problem of distributional shift [25]. These advances provide information that leads us to augment GRU with (and for an NAR) frequency gating for separating low-frequency (slow-varying, strategic behavior) and high-frequency (abrupt, tactical actions) components to better pad both the drift and sudden anomalies at the campaign level.

Irrespective of the improvement, there are several loopholes. Multimodal datasets with temporal granularity unique to the phenomena of crowdfunding are still rare; much of the existing literature brings in the experiences of fake-news or financial transactions with alternative labels. Measures that are reported are usually based on average accuracy or AUC but do not include calibration and early-warning performance, which is also fundamental to intervention on the platform. Finally, multimodal, temporality explainability is underdeveloped; besides the attention maps, techniques that assign cross-modal discrepancies to tangible objects would be more helpful in moderating workflows.

III. Methodology

The proposed Temporal Dynamics Aware Multi-Modal Fraud Detection Framework (TDMM-FDF) combines the text, visual, and temporal features to detect fraudulent crowdfunding campaigns. The architecture consists of nine steps, namely, data collection and preprocessing, feature extraction, temporal modelling, cross-modal consistency validation, feature selection, classification with FG-GRU, and blockchain verification. Fig. 1 depicts the block diagram of a sequential workflow in which a preprocessing stage is carried out on the text and images obtained in the form of crowdfunding campaigns and the features are extracted and aligned, and HM4 is used to model the temporal patterns. Such multimodal representations are merged with each other and maximized and then given into the FG-GRU classifier. The final legitimate results are recorded in a ledger infrastructure based on blockchain to provide integrity and security.

A. Data Collection and Preprocessing

The data used in this research was obtained from the Kickstarter crowdfunding platform and was heterogeneous in multimodal form, that are needed to conduct effective fraud analysis. The dataset used is summarized in Table I. In each campaign example, there were organized and unstructured items, such as the project title, entire narrative descriptions, creator profiles, visual media, logs of updates and backer comments in a timely manner. The time-sensitive aspect of these fields allowed the modelling of both the static correlations, as well as dynamic behavioral routes that are usually characteristic of a fraudulent campaign.

Considering the noisiness and informality of user-generated Kickstarter contents, an elaborate pretext text processing pipeline was used so as to make sure semantic integrated before feature extraction. First of all, the extraneous HTML tags, links, emojis, and non-standard punctuation were eliminated, and lexical distortion was avoided. Any textual information was then normalized in two steps: 1) lowercasing all texts to token uniformity, and 2) stop word elimination to bury high-frequency and low-information words. This normalization can be formulated as in Eq. (1):

$$T' = Normalize(T) = Lowercase\,(RemoveStopwords(T)) \quad (1)$$

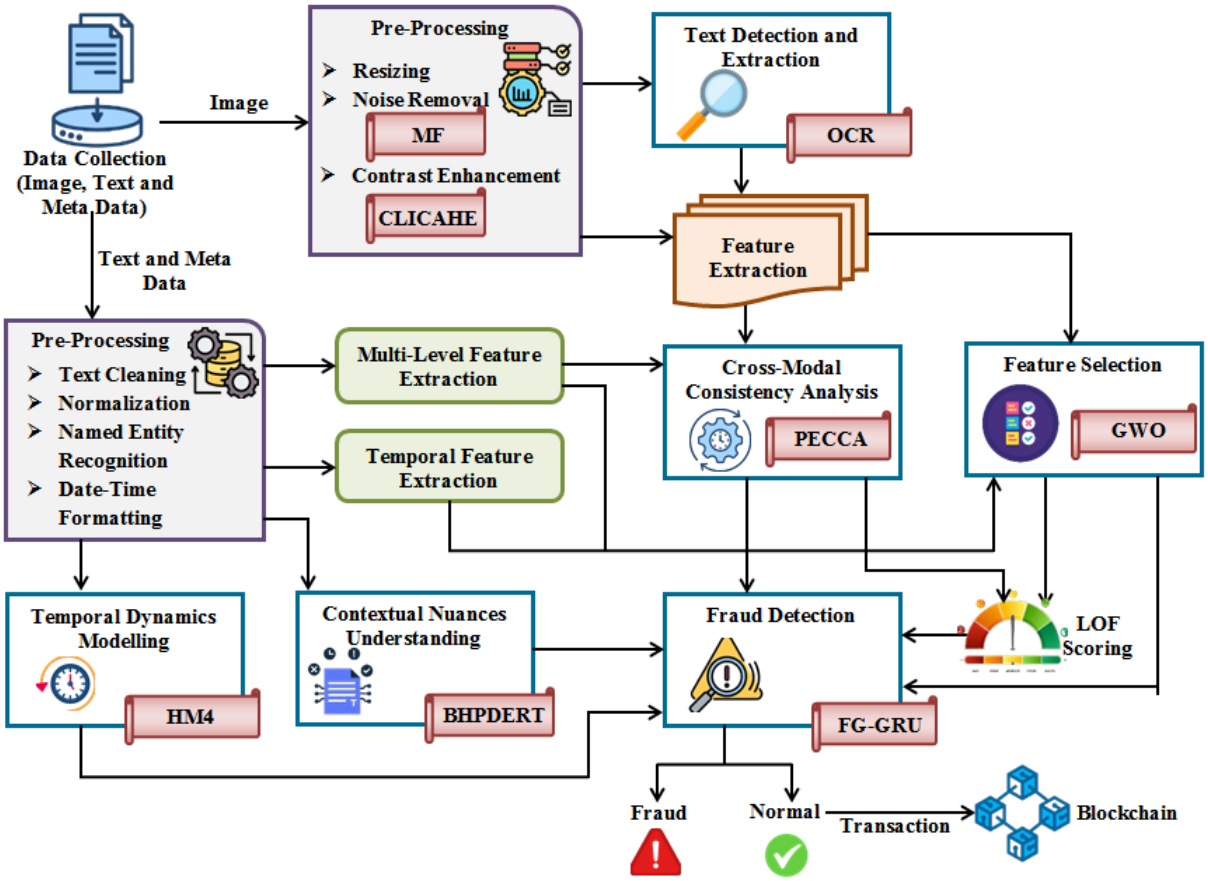

Fig. 1. Block diagram of the proposed framework.

TABLE I. DATASET AND PREPROCESSING SUMMARY

Attribute	Description
Number of campaigns	28,500 campaigns after preprocessing
Temporal coverage	2014–2024 Kickstarter data
Textual components	Titles, long descriptions, FAQs, updates, comments, creator metadata
Visual components	Poster images, prototype photos (1–8 per campaign)
Temporal sequences	Update timelines, backer activity logs, milestone timestamps
OCR-derived features	Extracted embedded poster text
Text preprocessing	HTML/emoji removal, normalization, stemming, NER
Image preprocessing	Median filtering, CLICAHE, OCR
Output modalities	Text vectors, image descriptors, temporal latent states

In addition, stemming was used to decrease inflexion forms, which minimized the inflexibility of token distribution. NER was later used to recognize semantically important entities which included organization names, product identifiers, time phrases, monetary values, and geopolitical positions. These entity-level annotations facilitated the verification of cross-image-based and time-based signals later on in the framework.

The image processing pipeline of the campaign was a combination of noise removal, texture-textuality-sensitive contrast segmentation, and inscribed text recovery to ensure the visual data quality brought across the actions of different campaigns matched adequately. Firstly, the images were morphologically enhanced with the Median Filtering (MF), which is a non-linear denoising operator that is efficient in noise reduction, especially salt-and-pepper noise, and the edges of the images are maintained. The process of filtering is given in the form of an equation, as shown in Eq. (2):

$$I'(x,y) = median\{I(i,j) \mid (i,j) \in N(x,y)\} \quad (2)$$

Subsequently, image contrast was improved using Contrast-Limited Image Complexity Adaptive Histogram Equalization (CLICAHE), an advanced extension of CLAHE. Unlike standard CLAHE, which applies a fixed clip limit, CLICAHE dynamically adjusts the clip threshold based on the Image Complexity Index (ICI) of each local region. Let H_k denote the histogram of a contextual region k, and let α_k [Eq. (3)] be the adaptive clip limit:

$$\alpha_k = \alpha_0(1 + \lambda \cdot ICI_k) \quad (3)$$

where, α_0 is the baseline clip limit, λ controls sensitivity to complexity, and ICI_k is defined as in Eq. (4):

$$ICI_k = \frac{1}{|R_k|}\sum_{(x,y)\in R_k}|\nabla I(x,y)| \quad (4)$$

Lastly, Optical Character Recognition (OCR) was used to extract textual clues that were within the poster-type images like product specifications, promotional statements, or disclaimers.

The OCR text was subsequently cross tabulated with the cleaned narrative descriptions to determine the presence of any textual variation or exaggerations that were suggestive of some form of fraud.

B. Textual Feature Extraction

The textual modality was enhanced by the extraction of lexical, semantic, and pragmatic cues, which make it possible to model linguistic behaviors of deceptive or misleading narratives in a comprehensive manner. The lexical indicators consisted of part-of-speech (POS) distributions, type-token ratios, and vocabulary richness measures, which all measure structural characteristics of campaign narratives. The pragmatic features have been calculated in order to distinguish between professionally written descriptions and suspiciously twisted or over-simplified text. Also, sentiment polarity and subjectivity scores were calculated to measure the tone of emotion, as fraudulent campaigns tend to be based on overstated optimism or framing of sentiment.

To have a deep semantic representation, the study utilized the Bidirectional Hierarchical Pattern Distillation Transformer (BERT-HPD), an improved transformer architecture, which aims to maintain the contextual depth and, at the same time, minimize computational costs. The underlying BERT encoder uses a contextualizing representation of all the tokens in the text by learning bi-directional dependencies between the text. Mathematically, given a sequence of input tokens ($w_1, w_2, \dots, w_n$), BERT will compute a contextual representation vector h_i [Eq. (5)] at every token position:

$$h_i = BERT(w_1, w_2, \dots, w_n) \tag{5}$$

As much as standard BERT offers high-quality semantic features, when operational on a large scale, its implementation can be expensive in terms of computation requirement to large multimodal pipelines. To solve this, a masked generation architecture (HPD) mechanism was incorporated that enabled the transfer of hierarchical linguistic knowledge of a high-capacity teacher BERT to a smaller student model. This distillation has a guarantee that significant syntactic/semantic/discourse-level structure is still present in the reduced representation. $h_l^{(teacher)}$ and $h_l^{(student)}$ will represent the hidden states of layer l of the teacher and student models, respectively. The objective of the HPD is to reduce the difference between representations on a layer basis and to be able to faithfully recreate linguistic hierarchical patterns. The loss during the distillation is determined as in the Eq. (6):

$$\mathcal{L}_{HPD} = \sum_{l=1}^{L} \left\| h_l^{(teacher)} - h_l^{(student)} \right\|_2^2 \tag{6}$$

The student model implements multi levels of linguistic abstraction, including local syntactic interactions as well as global semantic dependencies, without initially having the computational costs of the entire BERT architecture, by applying this constraint. The process does not only minimize the overfitting potential that would have occurred when, in the text training, it is important to train on text that has strong stylistic variations like crowdfunding campaigns but also retains fine-grained deception signals such as semantic inconsistencies, over-general content, and unnatural emphasis patterns. Consequently, BERT-HPD offers a high-capacity and effective basis of downstream multimodal fraud detection.

C. Temporal Dynamics Modelling

To model the temporal evolution of campaign behavior, this work employs the HM4, an advanced variant of the traditional HMM. HM4 is designed to capture long-range dependencies and structural dynamics in sequential update patterns that often characterize legitimate and fraudulent crowdfunding projects. Let the hidden state space be defined as, $S = \{s_1, s_2, \dots, s_k\}$, where each latent state corresponds to an underlying behavioural regime such as consistent updates, erratic communication, sudden activity surges, or prolonged inactivity. Similarly, the observable sequence is represented as, $O = \{o_1, o_2, \dots, o_T\}$, derived from timestamped campaign updates, linguistic tone shifts, and engagement metrics.

The classical HMMs have the transition matrix A, emission matrix B, and initial state distribution π, which are estimated by the Baum-Welch algorithm, which is an Expectation-Maximization (EM) implementation. Despite its popularity, Baum-Welch is iterative in nature and tends to become trapped in bad local minima especially when dealing with noisy or high-variability behavioral data like crowdfunding updates. This is the restriction that renders standard HMMs inadequate in detecting fraud cases when the underlying behavioral patterns are not close to stationary or smooth tracks.

To overcome these issues, HM4 is introduced to replace EM-based optimization with a global statistical moment-matching model, which allows analytical recovery of model parameters. The point is that when properly designed, low-order observable moments capture enough information concerning the dynamics of latent states. The first-order moment of observations meets the Eq. (7). Similarly, second-order cross-moments between adjacent observations capture transition structure:

$$E|O_t| = \sum_{i=1}^{k} \pi_i \mu_i, \qquad E|O_t O_{t+1}| = \pi_i A_{ij} \mu_i \mu_j^T \tag{7}$$

These moment equations constitute a web of algebraic equations. HM4 computes transmission and emission parameters without the need to find the transition matrix A or q in an iterative improvement HM4 computes the initial state distribution p, the emission parameters B and the overall transition matrix A by means of spectral decomposition or by means of the tensor's factorization directly. This does away with local minima vulnerability and offers a globally compatible forecast of the temporal dynamics.

D. Image Feature Extraction

After contrast enhancement and denoising, each processed image was then run through a detailed feature extraction pipeline that aimed to extract a set of complementary image features with respect to fraud detection. Images used in crowdfunding campaigns frequently include minor anomalies, including recycled or stock images, artificially-enhanced prototypes, or artificial visuals of products, that cannot be readily found by raw pixel inspection. To resolve this, a collection of well-defined handcrafted descriptors was used and each of them provided a different approach to texture, structure, and information in the key point level.

Local Binary Patterns (LBP) were first calculated to describe the patterns of micro-textures of the surface of the image. LBP represents local spatial difference by thresholding intensity of neighborhoods around each pixel thus producing rotation-invariant descriptors, which are very sensitive to material textures and surface consistency. This is especially concerning when trying to determine the inconsistencies on the surfaces of products or when trying to detect image patches which are artificial and are very common in fake campaigns. Then, Gray Level Co-occurrence Matrix (GLCM) characteristics were computed to measure the existence of higher-order spatial relations. GLCM is the intensity correlation of pairs of values at a set of offsets, and it can measure the contrast, homogeneity, entropy, and correlation. These characteristics give an understanding of structural coherence and can identify abnormalities like inappropriate lighting patterns, inappropriate shading, or a background that has been made up, which would suggest tampered or non-original images.

Histogram of Oriented Gradients (HOG) descriptors were applied to obtain geometric information. The distribution of gradient orientations is coded in HOG and the edges and contours and object boundaries can be represented in detail. This assists in revealing unnatural formation of edges or silhouettes that are overly smooth that can indicate image manipulation or utilization of unrealistic prototype images. The Scale-Invariant Feature Transform (SIFT) algorithm was also used to find strong key points and calculate local descriptions that are feature scale, rotation and illumination independent. SIFT is specifically good at recognizing recurring regions in images or recognizing the presence of an image borrowed by external, publicly accessible stock libraries. The resulting image-feature representation is an ensemble of all the descriptors obtained [Eq. (8)]:

$$F_{img} = [LBP, GLCM, HOG, SIFT] \quad (8)$$

E. Cross-Modal Consistency Check

The suggested framework includes the cross-modal consistency analysis mechanism, as PECCA, to be sure that textual and visual modalities are mutually consistent in terms of evidence. This module assesses the semantic consistency between campaign images and textual descriptions, which are essential in fraud detection since misleading campaigns tend to exhibit images that do not match with the textual description to give the illusion of being credible. Canonical Correlation Analysis (CCA) provides the classical foundation for Multiview alignment by learning linear projections of text features X and image features Y such that the correlation between their projected representations is maximized.

The architecture uses PECCA, which is a nonlinear extension of CCA that develops each modality by feature expansion via polynomials. These extended representations enable PECCA to represent the higher-order dependencies among modalities, which in effect capture subtle nonlinear dependencies, e.g., stylistic inconsistencies or unrealistic image-text co-occurrence or semantic discrepancies between the claimed product functionality and the visual representation. After its expansion, PECCA uses standard CCA on transformed feature spaces $\emptyset(X)$ and $\emptyset(Y)$, therefore integrating the analysis of nonlinear correlations and omitting deep learning-based fusion networks.

F. Feature Selection

After extracting text-based, image-based, and temporal features, the three modalities were concatenated [Eq. (9)] to form a unified multimodal representation. Whereby each element holds complementary data: the semantics of language and cues of deception in the text, structural and surface-level data in the images, and behavior patterns in the dynamics of time. This fused representation, despite being very expressive, generally falls into a high-dimensional feature space. This dimensionality may bring about redundancy, higher computational cost, and even diminished model generalization because of irrelevant or noisy attributes. Metaheuristic strategy of feature selection, which relies on the optimization of the grey wolf (GWO), was used to solve these problems.

$$F = [F_{text}, F_{img}, F_{temp}] \quad (9)$$

GWO is a bio-inspired optimization method which imitates the hierarchical leadership and hunter-cooperative behavior of the grey wolves. This hierarchy places the three most fit wolves in the form of α, β, and δ. They are the candidate solutions that are most fit in the feature space, and direct the search means the feature space. The rest of the wolves, the ω wolves, refresh their locations in accordance with such three leaders, and in this way, we have a balance in exploration and exploitation in the optimization. A fitness function based on classification accuracy was used to assess the quality of an individual candidate feature subset. In particular, a lightweight classifier was trained on a validation split based on the selected subset of features only, and the accuracy obtained was used as the fitness score.

G. Fraud Detection Model (FG-GRU)

The last prediction step performed by the proposed framework uses a Frequency-Gated Gated Recurrent Unit (FG-GRU), a recurrent network architecture specific to both gradual and rapid behavioral change, as well as sudden, high-intensity change that often indicates a fraudulent campaign. Compared to the classical GRU, which interacts dependency with time using its update and reset gates, the FG-GRU proves this operation, dividing the dynamics of the hidden state into low-frequency and high-frequency components and is able to significantly differentiate two long-term behavioral patterns and discrete upsurges. The base GRU computes its hidden state h_t at time step t using the gated recurrence formulation [Eq. (10)].

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tanh(Wx_t + r_t \odot Uh_{t-1}) \quad (10)$$

To better distinguish between smooth behavioral progressions and sharp deviations, the FG-GRU introduces a frequency decomposition stage. The hidden state h_t is passed through two filtering operators $F_{\text{low}}(\cdot)$, which extracts low-frequency (slow-varying) temporal components, and $F_{\text{high}}(\cdot)$, which extracts high-frequency (rapid-change) components. To integrate both frequency bands, FG-GRU introduces two learnable gating functions, g_{LF} and g_{HF}, which determine how much each component contributes to the final blended representation [Eq. (11)].

$$h_t^{FG} = g_{LF} \odot h_t^{LF} + g_{HF} \odot h_t^{HF} \quad (11)$$

To increase the reliability of the decisions, the FG-GRU output is assessed based on Local Outlier Factor (LOF). LOF

measures the isolation of a particular instance against the local neighborhood of such an instance in the latent representation space. Campaigns with multimodal temporal signatures that are significantly different than normal behavioral clusters are assigned lower density scores, and thus the prediction ambiguity is minimized, and more effectively distinguish between legitimate and fraudulent activities.

H. Blockchain Execution Layer

The framework uses a blockchain-based execution layer to make sure that the results of detecting the fraud are stored in a way that is tamper-resistant and auditable. This section reserves the results of the classification engine, namely, the campaign name, fraud rating, and final decision tag, into a decentralized registry. Using blockchain to anchor these findings will create an unalterable and transparent history of all the fraud evaluations conducted, which will instill greater confidence in the administrators of these platforms, campaign creators and sponsors.

A blockchain transaction Tx_i is created with each campaign C_i that has been taken through the multimodal evaluation pipeline. This is a transaction that includes key metadata, such as campaign ID, an output of the metadata, the fraud score S_i, or classification, and the time of evaluation t_i. The content of the transaction is encrypted with the help of a cryptographic hash, $H(\cdot)$, the implementation of which makes it impossible to make any changes to the stored information.

After it has been created, every transaction is added to the shared blockchain list stored by numerous nodes. The decentralization of the ledger makes it so that no one can manipulate or overwrite the outcome of a fraud assessment, and in the highly stakes setting of a crowdfunding site, where a challenge to the legitimacy of a campaign can emerge, it is of utmost importance. Additionally, the append-only cryptographically secured ledger does offer a verifiable audit trace that can be consumed either to conduct compliance audits, to resolve disputes, or to provide a long-term audit of campaign behavior.

IV. Results and Analysis

A. Simulation Setup

The study was performed on a real-life dataset that was gathered on Kickstarter, which is one of the biggest reward-based crowdfunding platforms. The structure of the dataset consists of campaign descriptions, project titles, visual media, creator metadata, update logs, timestamps, and funding progress indicators. Fraud labels have been generated using cases of scams reported publicly, campaigns identified as suspicious by the platform, and those reported by the community as fraudulent. To avoid the effect of temporal leakage, an inherent problem in sequential or time-dependent data where information in a future sample somehow affects the model's knowledge of historical behavior, the data has been stratified on a chronological basis. This guarantees that the time sequence of the campaign events is maintained during training and assessment. Namely, the first 70 per cent of campaigns were designated to the training set, which enabled the model to learn the patterns based on the historical data exclusively. The mid-period time frame entailed the subsequent 15% of campaigns, which made up the validation set and was utilized to tune hyperparameters and refine models. The last 15 percent of campaigns were used as the test set, which allowed the fair assessment of the model to generalize to new, unobservable campaigns that emerge in the future. Table II lists the most important hyperparameters for each component: HM4, PECCA, feature extractor modules, and the FG-GRU classifier.

TABLE II. Model Hyperparameters

Module	Parameter	Value / Description
HM4 (Temporal Model)	Hidden States (K)	6
	Moment Order Used	Up to 3rd-order cross-moments
	Transition Estimation	Algebraic MoM solver
	Emission Model	Gaussian Mixture (3 components)
Textual Encoder (BERT-HPD)	Max Seq Length	256 tokens
	Teacher Model	BERT-base (110M parameters)
	Student Model	8-layer distilled variant
	Distillation Loss	Layer-wise MSE + KL divergence
Image Processing	MF Kernel Size	5×5
	CLICAHE Tiles	8×8 blocks
	Clip-Limit	Adaptive (0.5–3.0)
Image Feature Extraction	LBP Radius	1
	GLCM Angles	{0°, 45°, 90°, 135°}
	HOG Cells	8×8
	SIFT Keypoints	Up to 500 per image
PECCA (Cross-Modal)	Polynomial Order	3
	Latent Dim (CCA Space)	120
	Correlation Threshold	0.55
FG-GRU Classifier	GRU Units	256
	Frequency Filters	Low (0–2 Hz), High (2–20 Hz)
	Batch Size	32
	Optimizer	AdamW
	Learning Rate	1.00E-04
	Epochs	50

B. Performance Comparison

To rigorously assess the efficiency of the developed multimodal fraud detector, the performance of the framework was compared to a collection of more popular baseline models, including classical ML methods, such as SVM and RF, and DL models, such as LSTM, GRU, and fine-tuned BERT. There were the same baselines train on the same experimental conditions, with the same chronologically stratified splits, and with the same normalized feature representations, in order to have an unbiased and fair comparison.

The outcomes of the performance, which are provided in Table III, indicate a steady and significant percentage of

improvement provided by the proposed model in all significant evaluation indicators. Conventional baselines like SVM and RF are also somewhat effective, but they fail to deal with the high temporal, visual, linguistic, and disparate nature of fraudulent crowdfunding activities. The sequential models, including LSTM and GRU, exhibit significant improvements since they can learn temporal dependencies; nevertheless, they cannot work well in the situation when cross-modal contradictions or irregularities in the behavioral pattern are present. The performance of Fine-tuned BERT is as good as it should be, especially regarding its contextual semantic knowledge, but it is restricted by its lack of inter-modal footing and explicitly modelled time view.

TABLE III. PERFORMANCE COMPARISON WITH BASELINE METHODS

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC	ECE
SVM	82.40%	79.50%	74.20%	76.70%	0.84	0.19
Random Forest	86.10%	83.80%	81.20%	82.50%	0.88	0.15
LSTM	88.70%	86.50%	84.10%	85.30%	0.9	0.13
GRU	90.30%	88.10%	85.60%	86.80%	0.92	0.12
BERT (Fine-tuned)	92.80%	91.20%	89.60%	90.40%	0.95	0.11
Proposed HM⁴ + PECCA + FG-GRU	96.40%	95.10%	94.20%	94.60%	0.982	0.06

The developed HM4 + PECCA + FG-GRU fusion architecture is obviously more superior. The system learns a much more diverse range of behavioral and semantic cues by combining HM4 with its strong ability to estimate temporal states, PECCA with its ability to model nonlinearly across cross-modal features, and frequency-aware sequential modelling, which is provided by FG-GRU. This leads to particularly high increases in Recall, a vital measure on fraud detection, and missing a fraudulent campaign is risky. The fact that the ROC-AUC score is very close to 0.98 is also a testimony to the high discriminatory power of the model in ranking suspicious campaigns over legitimate ones, suggesting that there is reliable discriminatory power even in difficult borderline cases. The other important benefit is the fact that the Expected Calibration Error (ECE) is reduced drastically between 0.11 with fine-tuned BERT and 0.06 in the offered system. This enhancement also indicates more credible confidence estimates, so that forecasts of fraud are closer to the likely outcomes, a necessary quality of real-world decision-making systems employed by crowdfunding platforms.

The proposed system has an ROC curve (Fig. 2) that shows a steep and very steep climb up to the upper-left corner, indicating that it has a high capacity to differentiate between fraudulent and legitimate campaigns at different decision thresholds. The model with a ROC-AUC of 0.982 has a significantly high discrimination capacity in comparison to all the baseline methods. This implies that, with the combination of time, text, and visual cues, the classifier is able to make accurate and correct ranking decisions even in borderline cases.

The Precision-Recall (PR) curve (Fig. 3) also proves the strength of the model to deal with the inherent class imbalance in fraud detection tasks. The suggested framework ensures that its precision exceeds 0.90 throughout a broad range of recall, which proves that the framework is capable of detecting fraudulent campaigns in an effective manner without the use of many false positives. This is especially crucial in real-life environments where a high rate of false alarms would destroy credibility and augment the operational load of investigation teams. The model proposed is highly precise at above 0.90 recall level, which can be explained by the fact that PECCA and FG-GRU complement each other with nonlinear multimodal alignment and frequency-conscious temporal modeling, respectively. These mechanisms taken jointly can provide a more detailed insight into campaign behavior and the system would hold true even in challenging conditions.

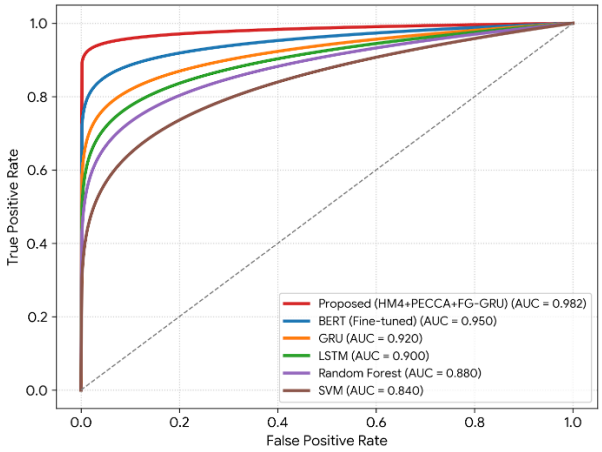

Fig. 2. ROC curve of the proposed fraud detection model.

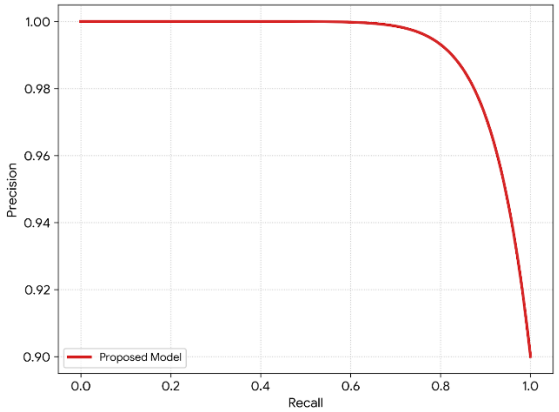

Fig. 3. Precision–Recall (PR) curve of the proposed fraud detection model.

As in Table IV, the confusion matrix obtained after testing the set shows that the proposed model possesses significantly low misclassification rates. Among all valid campaigns, 41 were falsely reported as fraudulent and the number of false negatives was only 23, which is significantly less than the baseline models. This enhancement underscores the usefulness of incorporating temporal state transitions of HM4 and multimodal contradiction detection of PECCA, which helps the system to identify subtle behavioural abnormalities that most simple models fail to detect.

Moreover, in the calibration curve (Fig. 4), it can be seen that there is a strong agreement between the predicted probability of fraud and the actual outcome frequency. It means that the model is effective in classification, as well as yielding credible confidence estimates. These predictability-aware predictions prove important at work with fraud-monitors, where the tolerance to operational risk or policy-specific to platform alert thresholds can be changed dynamically.

TABLE IV. CONFUSION MATRIX

	Predicted Legitimate	**Predicted Fraudulent**
Actual Legitimate	1035	41
Actual Fraudulent	23	412

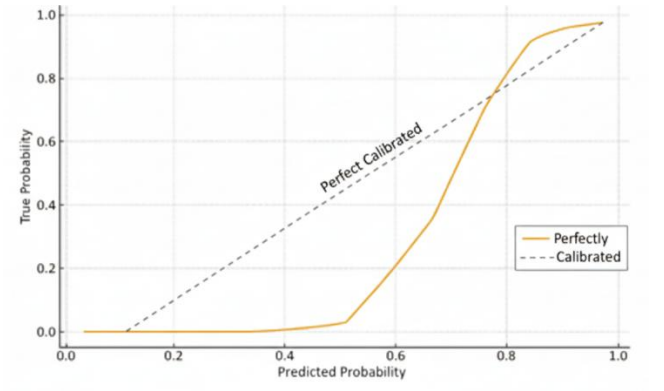

Fig. 4. Calibration curve of the proposed fraud detection

C. Computational Cost Analysis

To analyze three important aspects of computation, the runtime benchmarked three main aspects of computation, including training time per epoch, inference latency per sample, and memory usage on the GPU. The suggested model has greater computational overhead than the baseline GRU and fine-tuned BERT models, as illustrated in Table V. It has a 13.8 seconds per epoch training time and a 16.7ms per sample inference latency that indicates the extra processing added by multimodal feature integration, temporal moment modelling and frequency-gated recurrence. There is also the increment in GPU memory requirement to 7.9 GB due to the extended architecture and poly expansion modules.

A standard penciling of the blockchain integration in real-time detection systems is possible prohibitive latency, resource overhead. But in the TDMM-FDF architecture, the blockchain layer does not work as a processing module, but as an integrity-assurance layer, which is asynchronously running with the FG-GRU classifier. The system deters fraud scores (S_i) and metadata of campaigns by embedding them in a decentralized register, avoiding the so-called log-tampering that is a recognized vulnerability in a centralized platform management. We did a comparison of our blockchain-powered ledger and a regular centralized SQL-based logging system to determine that the integration is feasible. The findings, which are summarized in Table V, show that although the blockchain layer does add a slight increase in latency (22.4ms), the security advantages, namely, immutability and auditability by multiple parties, are essential to high-stakes governance of crowdfunding.

TABLE V. COMPUTATION COST ANALYSIS

Model	Training Time (per epoch)	Inference Latency (per sample)	GPU/System Memory
BERT (Fine-tuned)	11.2 s	14.3ms	6.5 GB
GRU (Vanilla)	4.3 s	5.1ms	2.1 GB
Proposed TDMM-FDF	13.8 s	16.7ms	7.9 GB
Blockchain Module	N/A (Asynchronous)	+22.4ms (Latency)	Negligible (CPU-bound)
Total Integrated System	13.8 s	39.1ms	7.9 GB + Storage

The computational trade-off would be fair even in the light of the extra costs because the gains in detection accuracy, recall, calibration reliability and ranking performance are significant. The enhanced predictive robustness and cross-modal interpretability of the suggested system offset the resource consumption increase in safety-critical settings, like fraud detection, where the outcomes of the failure to notice a fraudulent campaign can vary greatly in terms of financial and reputational implications.

D. Comparative Analysis of Feature Selection Methods

The choice to use the Grey Wolf Optimization (GWO) in the feature selection stage was due to the fact that it has a better balance of exploration and exploitation over the traditional metaheuristics. In high-dimensional multimodal spaces, such as that created by concatenating text, image and time features, approaches such as Genetic Algorithms (GA) are usually susceptible to premature convergence, whereas Particle Swarm Optimization (PSO) might get lost in local optima when optimizing non-convex fitness landscapes, as is the case with fraud detection data. In order to empirically defend this selection, a comparison was carried out with the use of a Kickstarter validation set. We compared GWO to GA, PSO and a standard L1-based (Lasso) selection. As the results, summarized in Table VI, indicate, GWO was able to find a smaller feature subset and a higher classification rate, which confirms its effectiveness in eliminating redundancy, and no important cues on deception were lost.

The GWO strategy achieved the largest reduction ratio (82.9) and the best fitness score (96.4), which was effective in recovering the curse of dimensionality, which is present in our 28,500 campaign data. This effectiveness is owed to alpha, beta, and delta leadership structure, which guides the search process towards the most informative region of the feature space faster as compared to the stochastic ones.

TABLE VI. COMPARATIVE ANALYSIS OF FEATURE SELECTION METHODS

Selection Method	Feature Reduction Ratio	Fitness (Acc)	Convergence (Iter)	Stability (Std Dev)
L1-Regularization	62.40%	91.80%	N/A (Linear)	±0.05
Genetic Algorithm (GA)	74.10%	93.50%	85	±0.12
Particle Swarm (PSO)	78.50%	94.20%	62	±0.09
Proposed GWO	82.90%	96.40%	48	±0.04

E. Ablation Study

The ablation study (Table VII) clearly demonstrates the contribution to the overall system performance of each of the components of the proposed architecture. The complete model scores the highest in all measures, proving the complementary advantages of the HM4, PECCA, BHPD-enhanced semantics, and the FG-GRU. The removal of the BHPD semantic distillation layer leads to a perceptible reduction in performance, which can be explained by the fact that the system would be less capable of detecting subtle linguistic information that is associated with campaign lie stories. The state-of-the-art HM4 leads to some of the greatest decreases in ROC-AUC and F1-Score, which shows that modeling the temporal state is important. The system cannot detect drift in behavioral patterns over time in the absence of HM4, and this increases the FNR. The removal of PECCA also harms the performance, especially the precision, because the model will be less effective in detecting the discrepancies between textual assertions and visual pieces of information.

TABLE VII. ABLATION STUDY RESULTS

Model Variant	Accuracy	F1-Score	ROC-AUC	Observation
Full Model (HM4 + PECCA + FG-GRU)	96.40%	94.60%	0.982	Best performance
Without BHPD (raw BERT)	94.10%	92.30%	0.964	Loss in semantic quality and deception cues
Without HM4	92.80%	90.70%	0.951	Temporal drift ignored; higher false negatives
Without PECCA	93.40%	91.10%	0.944	Missed text–image contradictions
Without FG-Gating (vanilla GRU)	91.90%	89.40%	0.935	Short-term spikes and long-term drifts poorly separated

Likewise, by substituting FG-GRU with regular GRU, the characteristics of the system to differentiate between the trends in long-term behavioral changes and unexpected anomalies decrease. This causes a poor depiction of irregularities in time, which in most cases are reflective of fraudulent activity. In general, the ablation findings indicate that all the components, that is, HM4, PECCA, BHPD, and FG-Gating, play a distinct and significant role in the robustness of the model, with HM4 leading the way toward Recall improvement, PECCA Precision improvement, and FG-GRU stabilization of the temporal dynamics through frequency-aware decomposition.

F. Discussion and Interpretations

The experimental outcomes show that the TDMM-FDF model is highly effective in overcoming the shortcomings of the traditional, one-dimensional models, because it can reflect the multifaceted, dynamic character of contemporary crowdfunding deception. The high and comparable performance of the proposed model compared to the baselines, such as BERT and vanilla GRU, indicates that fraud is not a linguistic or behavioral indicator but an emerged value of the cross-modal inconsistencies over time. As an example, the high Recall (94.2) indicates that the frequency-aware gates in FG-GRU are especially skillful to find out the so-called strategic fraud, in which a generator may fabricate a semblance of legitimacy by using slow-varying trends at the expense of sudden anomalies in updates or interaction in order to deceive supporters.

Moreover, the capability of the PECCA module to highlight discrepancies between the project narratives and visual prototypes is a response to a serious weakness wherein, in this case, the scammers rely on the use of professional-grade imagery to conceal poor or plagiarized textual descriptions. In addition to the spectral nature of the transitions of latent states expected of an HM4, this cross-modal correspondence gives the HM4 a type of interpretability that DL models that are black boxes frequently lack. The framework provides platform moderators with practical information about a certain behavioral change and semantic-visual contradictions instead of binary labels.

Practically, such multi-dimensional assessments are resistant to tampering and auditable, as the integration of a blockchain-based execution layer would provide. This openness is essential to ensuring long-term confidence in automated moderation systems, as it leaves a history of evaluation that cannot be changed, that shields both honest creators and prospective supporters against arbitrary decisions by platforms. Finally, the computational overhead trade-off is compensated by a massive decrease in Expected Calibration Error (ECE) that is converted to more accurate and confident fraud forecasting in high-stakes financial settings.

V. CONCLUSION

This study has provided a detailed Temporal Dynamics Aware Multi-Modal Fraud Detection Framework (TDMM-FDF), which has been developed to work in a reward-based crowdfunding system wherein fraudulent actions are exhibited with intricate interplay of linguistic indicators, visual anomalies and temporal aberrations. The combination of the proposed HM4 spectral-temporal model, Polynomial Expansion Canonical Correlation Analysis (PECCA), and the Frequency-Gated GRU (FG-GRU) classifier allows the framework to show a strong capability to identify deceptive campaigns that unimodal or snapshot-based classifiers would otherwise not identify. The experimental findings prove the existence of a unique, complementary role of each of the modules to enhance the predictive ability of the system. HM4 is a reliable indicator of latent behavioral regimes and temporal drift patterns that have a strong predictive potential of manipulative activity. PECCA offers a logical system of bringing to light semantic and stylistic contradictions between the narrative and the images in the campaign- one of the most commonly misused features of

advanced scammers. Likewise, the frequency-aware decomposition of FG-GRU allows the model to be able to distinguish between slow-moving behavioral cues and abrupt anomalies, thus being able to model both strategy-level deception and bursts of manipulative behaviors in updates or engagement.

In addition to model performance, the incorporation of a blockchain-based execution layer implies that it provides fraud assessment logs with secure storage, tamper resilience, and auditability, therefore, facilitating platform governance and long-term trust in automated decision systems. Although predictive accuracy and multi-modes integration of the framework is very high, there are a number of limitations that need to be recognized. First, the existing TDMO-FDF framework uses a database that is mostly based on Kickstarter. Although this gives a huge sample of 28,500 campaigns, the results cannot be completely extrapolated to equity-based or donation-based platforms where fraud indicators and backer activity patterns are not that similar. Second, the PECCA is an effective approach to learn the nonlinear discrepancy but it does so by choosing a polynomial order, which might need manual adjustment to other platform architectures. Future research must enhance capable early-warning identification in most lingering campaigns, project explainability to multi-modal time thought, and device self-trained pre-training on supernumerary lofty unlabelled crowdfunding datasets. Further extensions can also add video partitioning, pioneered hop of creator histories, and proactive learning processions of real-world moderation.

REFERENCES

[1] S. Lee, W. Shafqat, and H.-C. Kim, "Backers Beware: Characteristics and Detection of Fraudulent crowdfunding Campaigns," *Sensors*, vol. 22, no. 19, p. 7677, Oct. 2022, doi: 10.3390/s22197677.

[2] E. Mollick, "The dynamics of crowdfunding: An exploratory study," *Journal of Business Venturing*, vol. 29, no. 1, pp. 1–16, Aug. 2013, doi: 10.1016/j.jbusvent.2013.06.005.

[3] B. Perez, S. R. Machado, J. T. A. Andrews, and N. Kourtellis, "I call BS: Fraud Detection in Crowdfunding Campaigns," *arXiv (Cornell University)*, Feb. 2022, doi: 10.48550/arxiv.2006.16849.

[4] S. Bernardino and J. F. Santos, "Crowdfunding: an exploratory study on knowledge, benefits and barriers perceived by young potential entrepreneurs," *Journal of Risk and Financial Management*, vol. 13, no. 4, p. 81, Apr. 2020, doi: 10.3390/jrfm13040081.

[5] M. Machado *et al.*, "Crowdfunding Fraud Detection: A Systematic Review Highlights AI and Blockchain using Topic Modeling," *SSRN Electronic Journal*, Jan. 2024, doi: 10.2139/ssrn.4948895.

[6] W. Hou and J. Qu, "BM5-SP-SC: a dual model architecture for contradiction detection on crowdfunding projects," *Current Applied Science and Technology*, vol. 23, no. 6, Apr. 2023, doi: 10.55003/cast.2023.06.23.007.

[7] L. F. Cardona, J. A. Guzmán-Luna, and J. A. Restrepo-Carmona, "Bibliometric analysis of the machine learning applications in fraud detection on crowdfunding platforms," *Journal of Risk and Financial Management*, vol. 17, no. 8, p. 352, Aug. 2024, doi: 10.3390/jrfm17080352.

[8] E. Solodoha, "How much is too much? The impact of update frequency on crowdfunding success," *Administrative Sciences*, vol. 14, no. 12, p. 324, Dec. 2024, doi: 10.3390/admsci14120324.

[9] Y. Zhu, T. Peng, S. Su, and C. Li, "Neighbor-consistent multi-modal canonical correlations for feature fusion," *Infrared Physics & Technology*, vol. 123, p. 104057, Jan. 2022, doi: 10.1016/j.infrared.2022.104057.

[10] S.-Y. Lin, Y.-C. Chen, Y.-H. Chang, S.-H. Lo, and K.-M. Chao, "Text–image multimodal fusion model for enhanced fake news detection," *Science Progress*, vol. 107, no. 4, p. 368504241292685, Oct. 2024, doi: 10.1177/00368504241292685.

[11] X. Shen, M. Huang, Z. Hu, S. Cai, and T. Zhou, "Multimodal Fake News Detection with Contrastive Learning and Optimal Transport," *Frontiers in Computer Science*, vol. 6, Nov. 2024, doi: 10.3389/fcomp.2024.1473457.

[12] J. Hua, X. Cui, X. Li, K. Tang, and P. Zhu, "Multimodal fake news detection through data augmentation-based contrastive learning," *Applied Soft Computing*, vol. 136, p. 110125, Feb. 2023, doi: 10.1016/j.asoc.2023.110125.

[13] M. Nasser *et al.*, "A systematic review of multimodal fake news detection on social media using deep learning models," *Results in Engineering*, vol. 26, p. 104752, Apr. 2025, doi: 10.1016/j.rineng.2025.104752.

[14] Y. Shen, Q. Liu, N. Guo, J. Yuan, and Y. Yang, "Fake news detection on social networks: a survey," *Applied Sciences*, vol. 13, no. 21, p. 11877, Oct. 2023, doi: 10.3390/app132111877.

[15] V. M, J. Mohan, S. S. Kumar, and K. P. Soman, "Synergistic detection of multimodal fake news leveraging TextGCN and Vision Transformer," *Procedia Computer Science*, vol. 235, pp. 142–151, Jan. 2024, doi: 10.1016/j.procs.2024.04.017.

[16] S. Hangloo and B. Arora, "Feature fusion for multimodal fake news detection," *Procedia Computer Science*, vol. 259, pp. 1144–1153, Jan. 2025, doi: 10.1016/j.procs.2025.04.069.

[17] M. Huang, S. Jia, Z. Zhou, Y. Ju, J. Cai, and S. Lyu, "Exposing Text-Image inconsistency using diffusion models," *arXiv (Cornell University)*, Apr. 2024, doi: 10.48550/arxiv.2404.18033.

[18] S. Kumari and M. P. Singh, "A deep learning multimodal framework for fake news detection," *Engineering Technology & Applied Science Research*, vol. 14, no. 5, pp. 16527–16533, Oct. 2024, doi: 10.48084/etasr.8170.

[19] A. Adesh, S. G, J. Shetty, and L. Xu, "Local outlier factor for anomaly detection in HPCC systems," *Journal of Parallel and Distributed Computing*, vol. 192, p. 104923, May 2024, doi: 10.1016/j.jpdc.2024.104923.

[20] A. Cherif, A. Badhib, H. Ammar, S. Alshehri, M. Kalkatawi, and A. Imine, "Credit card fraud detection in the era of disruptive technologies: A systematic review," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 1, pp. 145–174, Dec. 2022, doi: 10.1016/j.jksuci.2022.11.008.

[21] A. R. Khalid, N. Owoh, O. Uthmani, M. Ashawa, J. Osamor, and J. Adejoh, "Enhancing Credit Card Fraud Detection: An ensemble Machine learning approach," *Big Data and Cognitive Computing*, vol. 8, no. 1, p. 6, Jan. 2024, doi: 10.3390/bdcc8010006.

[22] M. M. Ismail and M. A. Haq, "Enhancing enterprise financial fraud detection using machine learning," *Engineering Technology & Applied Science Research*, vol. 14, no. 4, pp. 14854–14861, Aug. 2024, doi: 10.48084/etasr.7437.

[23] X.-D. Liu, B.-H. Hou, Z.-J. Xie, N. Feng, and X.-P. Dong, "Integrating gated recurrent unit in graph neural network to improve infectious disease prediction: an attempt," *Frontiers in Public Health*, vol. 12, p. 1397260, May 2024, doi: 10.3389/fpubh.2024.1397260.

[24] C. Liu, F. Gao, Q. Zhao, and M. Zhang, "The optimized gate recurrent unit based on improved evolutionary algorithm to predict stock market returns," *RAIRO. Operations Research*, vol. 57, no. 2, pp. 743–759, Mar. 2023, doi: 10.1051/ro/2023029.

[25] X. Zhu, Y. Zhang, H. Ying, H. Chi, G. Sun, and L. Zeng, "Modeling epidemic dynamics using Graph Attention based Spatial Temporal networks," *PLoS ONE*, vol. 19, no. 7, p. e0307159, Jul. 2024, doi: 10.1371/journal.pone.0307159.