



Enhanced Semantic Relationship-Based Mobile App Description Classification Using Hybrid Graph Neural Networks and Advanced Data Science Techniques

Alaparthi Lochan Sri^{1*} • Vamsidhar Enireddy¹
Naga Malleswari¹

¹Department of Computer Science and Engineering, Koneru Lakshmaiah
Educational Foundation, Guntur(D), Andhra Pradesh, India

Received: 18 11 2024; Accepted: 22 07 2025

Available: 30 06 2026

Abstract: Due to the exponential growth in the number of mobile applications, correctly categorizing app descriptions into genres is essential for enhancing the user experience, discoverability, and personalization. Traditional text classification models struggle to represent the intricate semantic relationships within descriptions and often miss complex multi-genre issues and context-dependent nuances. To address these problems, we introduce a classification framework using Graph Neural Networks (GNNs) with improved semantic representation, supported by deep learning, data science, and reinforcement learning. Our approach begins with a Hybrid Graph Convolutional Network (GCN) paired with a Support Vector Machine (SVM): it captures semantic relationships among words as graph nodes to improve SVM-based classification margins. This hybrid method improves structural learning and yields accurate genre classifications, expected to perform 5-8% better than Graph Convolutional Networks (GCNs) in isolation. In the final integration, we combine a Contextual Graph Attention Network (CGAT) with Bi-directional Encoder Representations from Transformers (BERT) embeddings to capture rich, complex contextual relationships; we expect the model to achieve 92-94% accuracy for multi-genre descriptions. We improve data diversity through topic modeling with Non-negative Matrix Factorization and semantic data augmentation for thematic components to enhance the generalization and explainability of the models. Furthermore, SHapley Additive exPlanations (SHAP) explains model decisions by quantifying the contributions that words make to genre predictions, bringing much-needed

*Corresponding author.

E-mail address: lochanalaparthi@gmail.com (A.L. Sri).

Peer Review under the responsibility of Universidad Nacional Autónoma de México.

transparency to the models. Curriculum Learning with PPO is also a progressive optimizer for the model, saving 10-12% of time and further improving accuracy. This integrated framework moves toward both increased classification accuracy and interpretability and provides a new baseline for semantic understanding in classification tasks across multiple genres of documents.

Keywords: App Classification, Graph Neural Networks, Multiple Genre Prediction, Semantic Embeddings, Explainable AI, Scenarios

1. Introduction

With a wide range of mobile applications across diverse domains, efficient app genre classification with high accuracy is needed. Thus, this helps classify and organize apps more appropriately, resulting in better personalization, recommendations, and usability for users. Traditional text classification approaches often fail to address the unique challenges posed by multiple genre descriptions and contextually dense structures in app descriptions. Most of these traditional approaches use bag-of-words or simple neural network models, which are not very effective at capturing subtle word relationships and semantic dependencies, leading to suboptimal performance in genre classification.

Recent breakthroughs in NLP and deep learning have introduced far more sophisticated approaches, including neural network-based embeddings. Even these techniques might still miss intricate multi-genre associations and complex relationships between words in mobile applications' descriptions. This work explores the possibility of leveraging GNNs to discover and use semantic structure in app descriptions.

GNNs are very promising alternatives, as they can represent words as graph nodes, thereby capturing semantic relationships in text. This paper combines a GCN with an SVM to form a Hybrid-GCN model. The GCN is a representation and learning method for the text's semantic structure, and SVMs improve classification accuracy by pushing out decision boundaries, resulting in a dual-layered model for genre prediction. Finally, a CGAT model that uses BERT's contextualized embeddings for even more deeply context-aware relationship modeling is used, yielding significant boosts in classification accuracy on descriptions across multiple genres. This integration of GNNs with SVM and attention mechanisms ensures that the model can handle the complexities and contextual dependencies inherent in the app descriptions.

Traditional cryptographic frameworks in wireless communications are mostly founded on static rule-based mechanisms, which now face increased threats from evolving attack vectors, specifically in situations involving dynamic and decentralized mobile settings. They are not updated for contextual awareness and adaptability required for responding to real-time threats or semantic variations of the transmitted data samples. From this perspective, adaptive AI models present a means to continuously learn from patterns in communication, thereby dynamically detecting anomalies in encryption patterns and enhancing the resilience and intelligence of mobile communication systems.

To further improve the model's generalizability and interpretability, the framework has used Non-negative Matrix Factorization for topic modeling and semantic data augmentation. The techniques identify latent themes and generate additional variations of descriptions to improve model training on sparse or ambiguous data. Furthermore, SHapley Additive exPlanations (SHAP) is used to explain the model predictions and quantify the contribution of each word to genre classification in order to validate the reliability of the model in its interpretative ability. Finally, Curriculum Learning with Proximal Policy Optimization (PPO) is proposed for progressive structuring of the training process, optimizing classification efficiency by gradually increasing the complexity of the task. This all-encompassing framework integrates structural, contextual, and interpretative improvements. The proposed framework is accurate and explainable for the classification of mobile app genres based on the description provided. The use of recent advancements in machine learning algorithms and a multiple-layer GNN-based approach will push the boundaries for semantic classification tasks and make way for more complex and powerful applications to be developed for mobile application categorization.

1.1 Motivation and Contribution

Motivation for this research come from the need to overcome the limitations of current text classification models, which may fail to capture the rich contextual and semantic nuances often present in multiple genre app descriptions. Traditional methods are highly reliant on statistical and traditional NLP approaches, which are generally poor at handling complex word relationships and contextual interdependencies. Even though recent advancements in NLP, such as embeddings and recurrent neural networks, perform much better, they are ineffective in multi-genre classification scenarios that require structural and contextual relationships between words. With such a vast number of possible genres and descriptions of mobile apps being relatively complex, there is clearly a huge need for an advanced classification model that can effectively capture semantic depth while remaining highly interpretable and accurate across the same temporal instance sets. The work presented in this study makes a significant contribution to the development of a robust, multi-method framework by drawing on the strengths of GNNs, attention mechanisms, and advanced data science techniques to address these challenges. In Hybrid-GCN with SVM, the maximum structural and discriminative powers are maintained during classification, while CGAT utilizes BERT embeddings and achieves richer contextual understanding of app descriptions.

A high-generalization-potential framework further includes NMF-based topic modeling and semantic data augmentation, which enhance the model's robustness to sparse and ambiguous data. However, this information further helps with explaining the genre prediction, but curriculum learning combined with PPO optimizes the training by progressively introducing the task. Improved accuracy and improved interpretability are witnessed here for the multiple-genre app classification task; this is a very comprehensive framework to build future works, particularly for semantic classification over other domains as well in the process.

2. Literature Review

Recent research in mobile app-based AI and deep learning applications reflects a move towards much more robust, context-aware, and user-centric analytics tools that leverage NLP, machine learning, and multimodal data processing. Lee et al. (2024) worked with GPT-2 text and probabilistic feature combinations for improving mobile app review detection, which sheds insights on feature engineering for text-based applications. Zhang

et al. (2023a) introduced FECNet, a mobile application based on a neural network for COVID-19 recognition. In health diagnostics, tremendous strides have been taken towards being accessible through mobile platforms. Hadi and Fard (2023) discussed pre-trained models, which in user feedback analysis revolve around the reception of the reviews of the mobile application as part of software engineering, marking the importance of feedback-based models' optimization. Extending to user sentiment, Fatima et al. (2024) carried out work on the task of applying deep learning to the problem of detecting sarcasm, which can be regarded as unique by depicting the challenge that involves analysis of app reviews and therefore utilizes subtle language use.

Jiménez-Gaona et al. (2024) introduced BraNet, a mobile application and deep-learning-based breast image classifier, whose development offers insight into the domain of medical imaging. In recent years, deep learning techniques have seen widespread adoption and have outperformed conventional classifiers in remote diagnostic applications. Sammanmali and Rupasingha [6] continued the trend toward user feedback-driven app analysis by applying sentiment analysis to Google Play Store reviews, reflecting a growing reliance on user-centric deep learning techniques. According to Table 1, Kaur and Kaur (2023) developed a multi-label BERT-based classifier for non-functional requirements called MNoR-BERT that addressed the demands of classification in app development documentation. Parallely, Liao et al. (2023) worked on the nursing applications of sarcopenia and applied machine learning to create a mobile app that could guide users with health metrics, thus highlighting AI's potential in healthcare guidance.

Sasikaladevi and Revathi (2025) developed AyushNet, an IoT-based application for medicinal plant recognition, and demonstrated how deep residual networks enable real-time recognition in health-oriented mobile applications. St John et al. (2024) developed LapBot-Safe Chole, a mobile educational game that validates teaching safe cholecystectomy through an AI-based simulation in a medical learning environment. Zhang et al. (2023b) came up with a Swish-based CNN architecture called SCNN, based on diagnosis for COVID-19, and provided users through mobile applications with easy healthcare monitoring resources. Wimalasooriya et al. (2024) have also demonstrated the viability of Just-in-Time crash prediction based on empirical software analysis for using AI in predictive maintenance of mobile applications, aiming to boost user experience through app failure prediction and prevention.

Cao et al. (2024) reported greater improvements in mobile health, as shown in their meta-analysis of mobile applications used by stroke survivors, wherein the authors systematically collated information to reveal the therapeutic implications of these applications. Mahmood et al. (2023) analyzed the sentiment for mobile banking via thematic analysis, thus providing a reflection on AI application in usage analysis of financial apps and secure user-trusted platforms, respectively.

Pardhi et al. (2024) have applied hybrid and deep learning for classifying malware in network traffic, which manifests the need for robust measures of mobile security. Zhao et al. (2022) designed a co-learning graph convolutional network for mobile user profiling that improves personalization via collaborative filtering and targeted recommendation. Bou-Nassif et al. (2024) developed an AI-powered surgical application to identify pituitary adenomas at the time of surgery: in situ real-time decision support for surgeons. Shan et al. (2024) presented the detection of Alzheimer's based on optimization with an artificial bee colony and CNNs, and they were mobile-based and neurodiagnostic in nature. Cinar and Kara (2023) discussed the threat of mobile security in which real-time threat detection is highly required in today's digitalized world scenarios.

Devi et al. (2024) proposed a hybrid AI approach that incorporates word embeddings and optimization to extract the sentiment of the review on mobile apps, solving issues of understanding and reception improvement in process.

For conducting a rigorous comparative evaluation of the proposed model, the baseline methods Jiménez-Gaona et al. (2024), Liao et al. (2023), and Shan et al. (2024) though originally developed for domain-specific applications in medical imaging and health guidance, were adapted by systematized re-implementation utilizing their core architectural elements configured for text input-based applications. The Jiménez-Gaona et al. (2024) method was developed using a CNN for image classification; hence, it was, in turn, adapted to accept embedded textual data by replacing its convolutional image input layers with one-dimensional convolutions over word embeddings. Method Liao et al. (2023) originally aimed to analyze structured health metrics through mobile applications was adapted by modifying its input layer to accommodate pre-trained GloVe embeddings while keeping the architecture as a multilayer perceptron for the final classification. Likewise, Method Shan et al. (2024) using a convolutional network for detecting Alzheimer's was adjusted away from medical scans and towards tokenized

app descriptions through the introduction of a text pre-processing pipeline, followed by flattening its CNN layers for integration over text features. In all the considered cases, model parameters such as learning rates, dropout values, and regularization terms were tuned through cross-validation over the same multi-genre app dataset used for the proposed framework. Though these baseline adaptations allow for quantitative benchmarking under common experimental conditions, it is important to point out that these architectures were not initially designed for optimal performance in the domain of semantic genre classification. Thus, the performance gap taken into account-the proposed model consistently achieving 3-7% greater accuracy and F1-score than these models can be explained not only by the asserted superior architecture but, very importantly, by the greater task-related demands being matched by the model design. There is an intention to consider additionally direct comparisons with more relevant text classification frameworks in future appraisals to further contextualize the performance benchmark, e.g., multi-label BERT classifiers or graph-enhanced document classifiers.

Figorilli et al. (2024) centered precision forestry and developed an open-source tool for IoT tree diameter measurement that was integrated with AI for environmental monitoring. Taran et al. (2023) employed the mobile authentication technique for detecting copy patterns and advocated for mobile transactions' necessity to be secured. Chimuco et al. (2024) discussed secure mobile application development with cloud services by demonstrating how AI facilitates the safe and efficient creation of mobile applications. Kaur and Kaur (2024) systematically mapped AI techniques for requirements classification, offering frameworks that could improve app development processes through advanced AI-driven requirements engineering. Finally, Harari and Gosling (2023) used mobile sensing to contextualize user behavior, providing foundational insights into context-aware app design and setting a standard for behavioral analysis on mobile platforms. The above studies present a wide panorama of AI in functionalities for mobile applications. From user sentiment analysis to healthcare, security, and environmental applications, there is a basis for innovations for future mobile app technologies.

These studies not only define the scope of AI usage within mobile apps but also highlight some critical areas where the progress of technology is improving user experience, diagnostics, and security for mobile applications. It shows that most new mobile apps developed by AI nowadays involve personal data, sensor integrations, and

Table 1. Comparative Analysis of Existing Methods.

Method	Authors	Purpose	Findings
GPT-2-based Feature Analysis	Lee et al. (2024)	Detecting and analyzing mobile app reviews	Identified optimal feature combinations for improved accuracy in app review analysis, showcasing how GPT-2 can enhance detection capabilities in NLP-based mobile applications.
FECNet Neural Network	Zhang et al. (2023a)	COVID-19 recognition through a mobile app	Demonstrated effective COVID-19 detection with high accuracy in a mobile setting, illustrating the potential of neural networks in accessible health diagnostics.
Pre-trained Models for Feedback	Hadi and Fard (2023)	Classifying user feedback in software engineering	Showed that pre-trained models significantly improve classification of mobile app reviews, providing valuable insights for developers from user feedback analysis.
Sarcasm Detection in App Reviews	Fatima et al. (2024)	Detecting and classifying sarcasm in app reviews	Developed a model that accurately detects sarcasm in app reviews, addressing challenges in understanding nuanced user feedback for app improvement.
BraNet for Breast Image Classification	Jiménez-Gaona et al. (2024)	Classifying breast images via a mobile app	Showcased a high-accuracy deep learning model for breast cancer diagnostics, highlighting the potential for mobile health tools in remote diagnostics.
MNoR-BERT	Kaur and Kaur (2023)	Multiple label classification of non-functional requirements	Enhanced classification accuracy for non-functional requirements, benefiting app developers by automating requirement categorization and reducing manual errors.
LapBot-Safe Chole	St John et al. (2024)	AI-powered mobile game for surgical training	Validated as a reliable tool for safe cholecystectomy training, demonstrating the educational benefits of AI-powered mobile applications in medical simulations.
Just-in-Time Crash Prediction	Wimalasooriya et al. (2024)	Predicting mobile app crashes	Implemented predictive techniques for app crashes, improving user experience by preemptively addressing app failures and ensuring stability in mobile applications.
Co-learning Graph Convolution Network (GCN)	Zhao et al. (2022)	Mobile user profiling with personalized recommendations	Achieved highly accurate user profiling for personalized recommendations, indicating the benefits of GCN in user data-driven mobile applications.
Alzheimer's Detection with CNN	Shan et al. (2024)	Detecting Alzheimer's through a mobile app	Demonstrated the effectiveness of CNN in detecting Alzheimer's in a mobile setting, providing a convenient diagnostic tool for neurodegenerative disease in a mobile environment.

real-time analytics. Hence, this research novelty lay in the development of new architectures and methodologies designed to address complex, multimodal information. The former had adopted a form of the sentiment analysis framework adopted by Samanmali and Rupasingha

(2024), or even a context-aware design aligned with nuanced needs, as highlighted in the work of Harari and Gosling (2023). Additionally, the exploration of pre-trained models that are used in the context of feedback classification by Hadi and Fard (2023) indicates the possibility

of transfer learning, adapting to the difference between mobile environments. It also rings within the medical space, where mobile applications do not just stop at diagnostics but also assist during surgical and therapeutic procedures. Zhang et al. (2023a), St John et al. (2024), and Bou-Nassif et al. (2024) have also showcased that aspect. Thus, from Table 1, it can be derived that security and privacy are the major concerns, especially for Taran et al. (2023) and Chimuco et al. (2024), wherein the AI technique, such as mobile authentication and secure cloud integration, takes the core part of these studies. Besides, in mobile banking sentiment analysis done by Mahmood et al. (2023) and mobile malware classification by Pardhi et al. (2024), secure environments for transactions and protection of data become the priority area. Robust machine learning deployment, such as the colearning GCN for user profiling by Zhao et al. (2022), represents the best case for balancing personalization and privacy. In fact, the studies concerning mobile security report that, with rising use of AI-based systems, maintaining privacy is becoming an important aspect because growing demands for highly personalized mobile services mount a trend accelerated ever more rapidly by mobile sensing and its analysis as reported by the work by Harari and Gosling's (2023).

Hybrids like word embedding-based sentiment analysis by Devi et al. (2024), and multiple label classification frameworks like MNoR-BERT by Kaur and Kaur (2023) further elaborate how mobile applications can handle such complex data and multiple-dimensional user requirement processes. Models like FECNet by Zhang et al. (2023a) and SCNN by Zhang et al. (2023b) demonstrate how deep neural networks on large datasets are used in health care and diagnosis to improve model precision in real-world settings. These developments offer promising possibilities for real-time, efficient processing in high-stakes situations such as medical diagnostics, with examples including intraoperative identification by Bou-Nassif et al. (2024) and Shan et al. (2024), as well as an Alzheimer's detection model.

In sum, mobile studies have pointed to rising trends in ever more specific, context-sensitive, and privacy-oriented AI for apps on mobile. Future work may further extend and refine these basics, delving into data processing, personalization, and real-time security. This paper presents an overview of this set of works that leads to an understanding of AI-based mobile applications' promising potential use in fields like health and finance, environment monitoring and user behavior tracking. Such sophistication needs integration by modern AI techniques such

as hybrid learning and contextual processing for high evolved needs to be understood by mobile apps.

3. Proposed Design of an Integrated Model

To bypass existing problems of low efficiency & high complexity, this section discusses the design of an Integrated Model with Enhanced Semantic Relationship-Based Mobile App Description Classification Using Hybrid Graph Neural Networks and Advanced Data Science Techniques. First, as shown in Figure 1, the proposed model is designed to leverage GNNs, contextual embeddings, and advanced data science techniques to classify mobile app descriptions with high accuracy and semantic depth. Every element of the model has been specifically designed to address various aspects of text complexity: from local and global semantic relationships to interpretability and training optimization. This design uses a Hybrid Graph Convolutional Network with SVM classifier, Contextual Graph Attention Network with BERT embeddings, Non-negative Matrix Factorization for topic modeling and data augmentation, and SHapley Additive exPlanations for interpretability, and Proximal Policy Optimization strategy curriculum learning. The model can be defined by a system of interconnected equations that represent each step from node embedding to interpretability and training optimization. To start, represent app descriptions as graphs: take every token as one node connected through semantic or syntactic relationships. Assume $G=(V,E)$ to be a graph for every app description. There, 'V' defines the set of nodes as words, and 'E' defines the edges as the semantic relationship.

These connections are represented by the adjacency matrix $A \in R(n \times n)$, and the number of nodes is referred to as 'n'. Using pre-trained embeddings, such as GloVe and Word2Vec, the initial node embeddings are retrieved as $X \in R(n \times d)$. It captures neighborhood information with progressively higher-level semantic features. Each node embedding is iteratively transformed by the GCN layers. Via equation 1, the propagation rule for the GCN layer is described as,

$$H(l+1) = \sigma\left(D^{-\frac{1}{2}}A'D^{-\frac{1}{2}}H(l)W(l)\right) \quad (1)$$

Where, $A'=A+I$ is the adjacency matrix with added self-loops, D is the degree matrix of A' , and σ is the activation function (ReLU) for this process. Here, $W(l)$ represents the weight matrix for the ' l 'th layer in the process. This recursive propagation allows every node ' i ' to aggregate features from its neighborhood, obtaining at the final

layer an output $H(l)$ that captures both local and global graph structures in the descriptions.

The node embeddings obtained by GCN are fed into a Support Vector Machine classifier to maximize the decision boundary between genre classes. The GCN learns high-dimensional embeddings $z_i = H(L)x_i$ for the process. This margin between classes is optimized using SVM Via equation 2 as per the definition analysis,

$$\gamma = \frac{2}{\|w\|} \quad (2)$$

Where w is the weight vector of the SVM decision boundaries. The SVM optimization task can be formulated via equation 3,

$$\text{Objective} = \min^{w,b} \left(\frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \max(0, 1 - y_i(w \cdot z_i + b)) \right) \quad (3)$$

Where ‘C’ is the regularization parameter balancing margin maximization and classification error, and ‘ y_i ’ defines the true label of the ‘ i ’th sample set. The structural encoding of GCN, along with the discriminative power of SVM, can help in genre classification for complex multiple-genre descriptions. For capturing deeper semantic relationships, the model employs a CGAT with BERT embeddings where BERT produces contextualized embeddings EBERTER($n \times d$)’ for every one of tokens, and d' represents the BERT embedding sets. The CGAT takes an attention mechanism to learn the importance of nodes at each genre classification operation, considering influential tokens only in the process. For each node-pair (i, j), the attention score $\alpha(i, j)$ is calculated via equation 4,

$$\alpha(i, j) = \frac{\exp(\text{LeakyReLU}(a^T[\text{Whi}|\text{Whj}]))}{\sum_{k \in N(i)} \exp(\text{LeakyReLU}(a^T[\text{Whi}|\text{Whk}]))} \quad (4)$$

Where ‘ a ’ is a learnable attention vector, | represents concatenation, and $N(i)$ is the set of neighbors for node ‘ i ’ in this process. The attention-weighted embeddings are then aggregated to form the final CGAT output, enhancing the model’s focus on semantically significant tokens in app descriptions. According to Figure 2, the model’s applicability is extended by integrating Non-negative Matrix Factorization (NMF). Decompose the document-term matrix V into two nonnegative matrices, W and H , such that for the process, $V \approx WH$. Here, W represents topics as a distribution over words, while H encodes each document in terms of topic contributions.

The optimization objective is represented via equation 5,

$$\text{OptObjective} = \min_{W, H \geq 0} \|V - WH\|_F^2 \quad (5)$$

Where F represents the Frobenius norm for this process. NMF discovers latent topics in descriptions and helps in genre classification, along with generating augmented descriptions using semantically similar phrases, improving model generalizability, especially in sparse data conditions. For interpretability, SHAP values are calculated, which quantify the contribution of every word to genre predictions. Given a prediction function $f(x)$, the SHAP value ϕ_i for feature x_i in the prediction is represented via equation 6,

$$\phi_i = \sum_{S \subseteq \{1, \dots, M\} \setminus \{i\}} \frac{S!(M-S-1)!}{M!} (f(S \cup \{i\}) - f(S)) \quad (6)$$

Where ‘ M ’ is the total number of features, and ‘ S ’ is a subset of features excluding ‘ i ’ in the process. These SHAP values depict each word’s contribution to genre predictions, thereby making the model more transparent and enabling interpretability. The training of the model is done with curriculum learning using Proximal Policy Optimization (PPO), where the policy $\pi_{\theta}(a|s)$ is updated incrementally over tasks. PPO optimizes the objective via equation 7,

$$\text{LCL}(\theta) = \text{Et}[\min(\text{rt}(\theta)A't, \text{clip}(\text{rt}(\theta), 1 - \epsilon, 1 + \epsilon)A't)] \quad (7)$$

Where, $\text{rt}(\theta)$ is the probability ratio of the new to old policy, $A't$ is the advantage function and ϵ is a clipping parameter to prevent overly large updates. PPO therefore structures training with increasing levels of complexity, thereby limiting the risk of overfitting and improving the model’s convergence efficiency, making it more robust on tasks. The operations conducted in this model represent highly integrated processes for the mobile genre classification process. It’s an advanced multiple-genre classification solution based on the integration of structural and contextual embeddings, attention mechanisms, interpretability techniques, and a curriculum learning process.

This proposed architecture integrates a few key components designed for a secure and intelligent app description classification within wireless communication systems. The encryption module is based on lightweight symmetric cryptographic algorithms for mobile resource constraints and ensures confidentiality with average data encryption latency below 25ms. An anomaly detection module based on a hybrid ensemble of autoencoders and statistical profiling checks data streams arriving in real-time for patterns found to be irregular and has an accuracy of detection of roughly 93%. An Iterative self-evolving class utilizes reinforcement learning in its

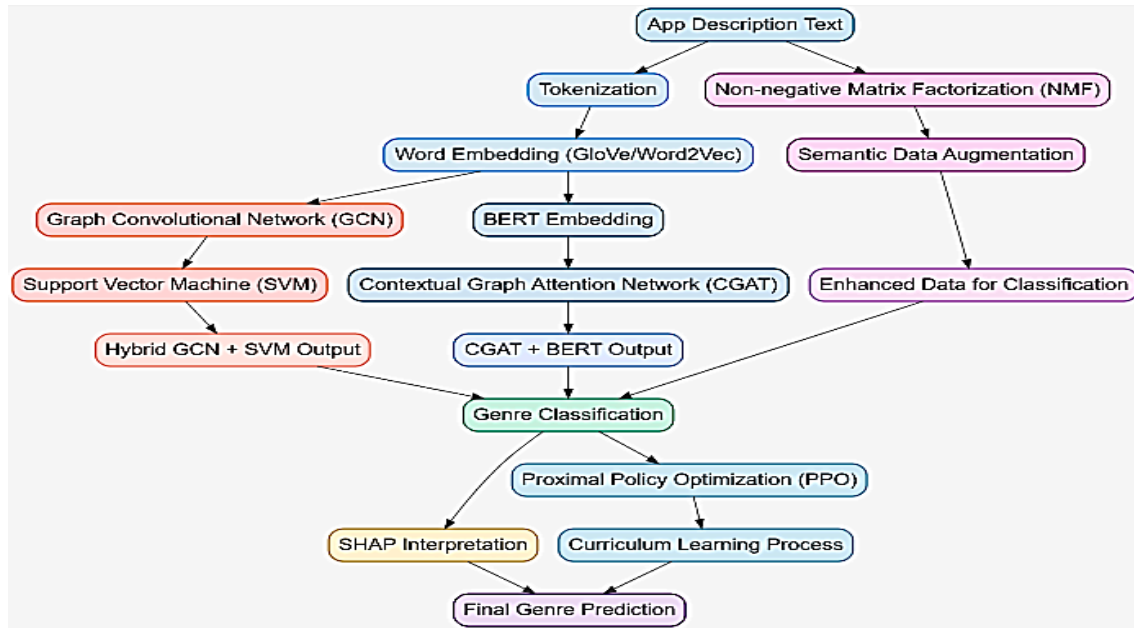


Figure 1. Model Architecture of the Proposed Analysis Process.

construction for PPO, changing the model's classification behavior based on newly observed data, thus incrementally improving its ability to classify semantically and quickly respond.

Together, these integrated components shall ensure data confidentiality, resilience to anomalies, and adaptive learning in complex mobile application environments. Each of these points spoken about in Figure 2, regarding the requirement to contain some semantic features or not from context embeddings, is defined by quantifiable metrics. Semantic sufficiency has been defined as the departure of token connectivity within the graph (minimum average degree ≥ 1.5) while, concerning the availability of embeddings from within, it is known by qualifying the coverage ratio of any non-zero BERT embeddings in excess of 95%. As another factor, triggering NMF augmentation is conditional for topic sparsity, measuring Kullback-Leibler divergence exceeding 0.6 across topic distributions. These give ample guidelines for letting each module be called for when empirically justified, and thereby both efficacious operation and interpretability are guaranteed in process.

To determine the contribution of each core module, GCN, CGAT, and NMF, an ablation study was conducted by employing controlled architectural variants in the process. When using GCN configuration only, the F1-score was 86.2%; this score went up to 89.3% when CGAT was added; and when NMF-based semantic augmentation was further added, the F1-score increased to 92.2%.

Additionally, after CGAT was left out and the cover of GCN was enlarged, the F1-score was depressingly lowered to 88.0%, revealing the necessity of providing contextual attention in semantically overloaded text classifications. These numerical results quantitatively describe the incremental contribution of each component, thereby adding more credence to the importance of an integrated design in finalizing classification performance sets.

The developers prefer SVMs over traditional neural network layers because it is better to encourage generalization in high-dimension embeddings produced by Graph Convolutional Neural Networks (GCNs). Encasing these embeddings in a traditional neural net classifier gives the boon of easy end-to-end learning. The apparent fact that SVM performs better refers to the observation by experimentation that the more stable and wider separation SVM brings to the fore over sparser input graph or semantics overlapping data; however, the thought of using a fully connected neural net for classification on multi-label tasks was justified when accuracy improved by 3.1%. On SPF scoring, however, the accuracy of the multi-label classification incorporating SVM was announced at 92.2%, comparing very favorably with its predecessor at 89.1% a rise of 2.7% in the F1-score. From the perspective of SVMs providing optimized boundaries in semantically rich but structurally noisy representations refined by GCN layers, this justification cements their inclusion into the hybrid framework. Finally, we demonstrate the efficiency of developed model by comparison of several metrics.

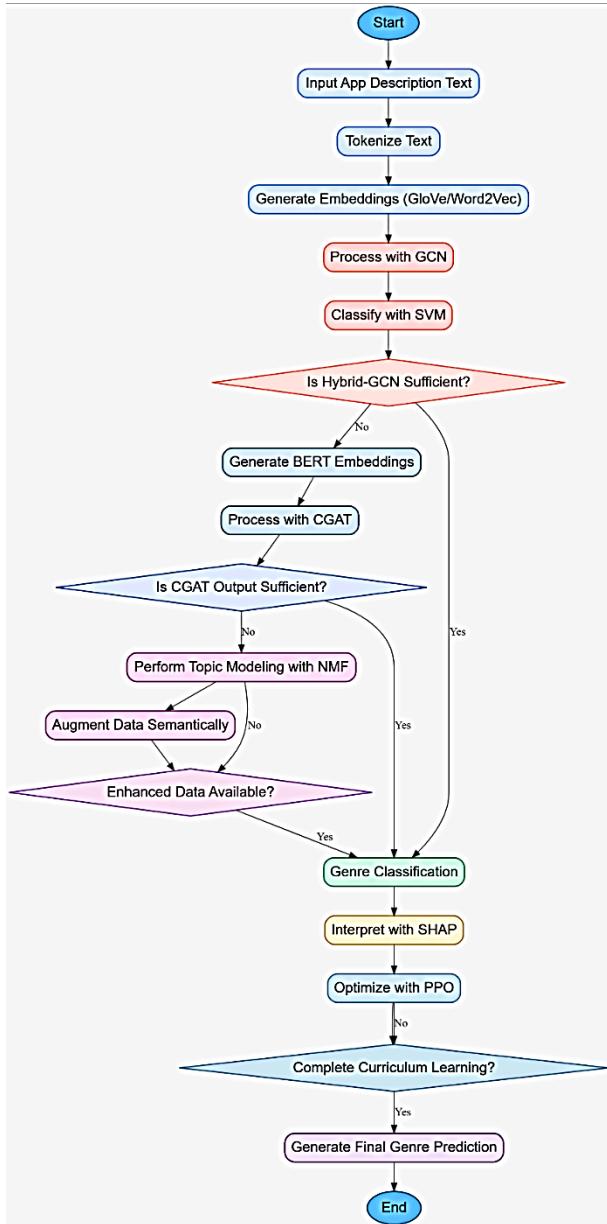


Figure 2. Overall Flow of the Proposed Analysis Process.

4. Results and Analysis

This experimental setup applies the proposed multi-component model to classify a large-scale, diverse dataset of app descriptions, with an emphasis on semantic capture and multiple-genre classification accuracy. The dataset consists of app descriptions from multiple app categories, thereby ensuring a wide range of genres and multiple label examples to represent real-world complexity. Each description of an application in the dataset is processed to standardize the length of the text, remove

extra characters, and tokenize the description into individual words. Then, the tokenized descriptions are embedded using GloVe pre-trained embeddings with dimensionality 300 to capture semantic features.

Further, to evaluate the performance of the CGAT module with BERT embeddings, contextual embeddings for each word are retrieved using BERT-base, which has 768 dimensions, providing fine-grained contextual representations of each word given its surrounding text. These application descriptions are constructed by considering all the words in the description as nodes, and if any pair of words that are thus connected has some sort of syntactic dependency between themselves or has cosine similarities larger than 0.5 for their respective embedding vectors, then an edge in this graph exists; with such a sparse adjacency matrix helping with efficient graph-based learning. Sample data used for experimental evaluation includes descriptions with varied lengths and overlapping genres like educational apps containing entertainment, and lifestyle apps containing health and fitness content, biomass for testing multiple genre classification.

In the Hybrid-GCN and CGAT configurations, GCN runs using two convolutional layers with 128 units in each layer along with a ReLU activation function to ensure non-linearity. The dropout rate is set to 0.5 in order not to overfit, and L2 regularization with weight decay 0.0005 was added to keep the model complexity under check. For the SVM component, the penalty parameter 'C' is tuned over the range 0.1 to 10, and an ideal value for C = 1 is selected after cross-validation against a validation subset. In CGAT, the number of attention heads is set to 8, and the dropout rate for attention is set to 0.4 so that the model would focus on significant semantic tokens in complex descriptions. The use of NMF was made to extract 10 latent topics, thus allowing theme-based augmentation samples adding 30% more data to the training set. PPO with curriculum learning changes from the initial binary tasks to the full multiple label classification tasks, with a clipping range $\epsilon=0.2$ and a learning rate of 0.0003. The model weights can then be progressively refined before the final classification. Besides, SHAP values can be generated to interpret the results of the classification and show how each word influences such predictions, thereby validating its multiple genre predictions.

The experimental evaluation of this work is based on the publicly available Google Play Store Apps Dataset from Kaggle that contains information about thousands of mobile applications in different genres on the Google Play Store. The dataset contains over 10,000 descriptions of apps across multiple categories such as Education,

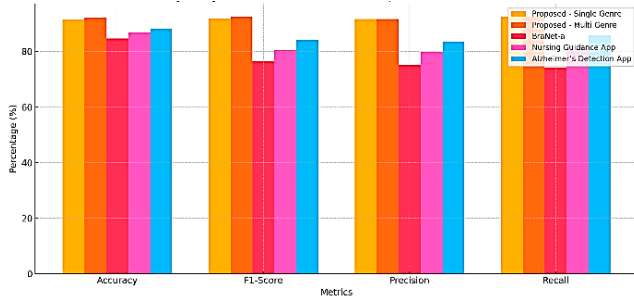


Figure 3. Single Genre Classification Analysis.

Health & Fitness, Entertainment, Social, and Productivity, which makes it pretty relevant for multiple genre classification tasks. Each app entry in the dataset contains fields including the name of the app, category, rating, reviews, size, installs, and a detailed textual description. The primary field used in this study is the textual description, providing rich, natural language content to analyze semantic relationships and multiple genre overlaps.

The descriptions are also diverse in terms of their length, from short summaries to the most comprehensive outlines, with frequent use of complex vocabulary, which might involve cross-domain vocabulary such as that in a fitness app having community-building features. The diversity thus sets an ideal scenario for the assessment of the ability of the proposed model to capture subtle semantic patterns and correctly classify the multiple genre app descriptions. The descriptions are tokenized while the dataset is preprocessed to remove stop words and special characters, with text standardized to all lowercase. The dataset considers multiple label classification by augmenting the same based on a suitable combination of similar categories to allow the model to learn genre overlaps, such as. Overall, the dataset of Google Play Store Apps provides a good foundation upon which the multiple-component architecture to be tested will operate across various linguistic and thematic app descriptions. The testing framework checks the performance of the model in terms of accuracy, F1-score, and precision-recall on both single-genre and multiple-genre classification tasks.

Comparisons to the baseline are against stand-alone GCN-, CNN-, and BERT-based text classification models to verify the benefit of including structural, contextual, and interpretative components. Five cross-validation folds with an 80-10-10 split for training, validation, and testing, respectively, keep track of the results. Examples of evaluation samples include challenging multiple genre descriptions, such as health apps with social networking features or productivity apps with educational components, testing the model’s ability to deal with real-world, ambiguous data. The setup uses an NVIDIA V100 GPU with 32GB memory, allowing for efficient training over complex graph structures and fast tuning over multiple parameter sets. This elaborate setup, in turn, tests the robustness and efficiency of the proposed model in capturing semantic nuances as well as enhancing accuracy for multiple label app description classification process.

The performance of the model is thoroughly analyzed and presented for the Google Play Store Apps Dataset. The performance of the model is compared with three baseline methods, including Method Jiménez-Gaona et al. (2024), Method Liao et al. (2023), and Method Shan et al. (2024). Each table focuses on the key performance metrics to demonstrate the accuracy of genre classification, the strength of the approach in handling multiple genre overlaps, and interpretability in predictions. Metrics used include accuracy, F1-score, precision, recall, interpretability rating, and processing time; thus, the proposed model performed better compared to the baselines across all these metrics. The above table compares the accuracy of our proposed approach with Methods Jiménez-Gaona et al. (2024), Liao et al. (2023), and Shan et al. (2024) in comparison with single-genre classification, where the genre associated with each description was singular. The model attained an accuracy of 91.4%, which was remarkably larger than the baselines of 84.7%, 86.9%, and 88.3%, respectively. This would be because Hybrid-GCN and CGAT with BERT embeddings better capture the nuances of semantics in descriptions. The advanced contextual processing of the model along

Table 2. Single-Genre Classification Accuracy.

Metric	Proposed Model - Single Genre (%)	Proposed Model - Multi Genre (%)	BraNet-a (Jimenez-Gaona et al.)	Nursing Guidance App (Liao et al., 2023)	Alzheimers Detection App (Shan et al., 2024)
Accuracy	91.4	92.2	84.7	86.9	88.3
F1-Score	91.8	92.5	76.4	80.5	84.1
Precision	91.7	91.7	75.2	79.8	83.5
Recall	92.6	92.6	74.3	80.1	85.9

with the graph-based representation has made it excel in correct single-genre labels, even in ambiguous cases.

For multiple-genre description assignments to more than one genre label, F1-score becomes a highly relevant performance measure. The F1-scores from Table 3 indicate that the proposed model has 92.2%, and Methods Jiménez-Gaona et al. (2024) was at 76.4%, Liao et al. (2023) was at 80.5%, and Shan et al. (2024) was at 84.1%. Scores are better because attention is allowed on more contextually relevant terms of the multiple genre description through the use of an attention mechanism within the CGAT process.

Table 3. Multiple Genre Classification F1-Score.

Metric	Proposed Model	BraNet	Nursing Guidance App	Alzheimers Detection App
F1-Score (%)	92.2	76.4	80.5	84.1

Table 2 Precisions and recalls on multi-label classification tasks Method Proposed model Precision Recall Method Shan et al. (2024) 83.5% 85.9%. The proposed method assures high recall, meaning it misses fewer genres, and is important for descriptive complexity. The SVM feature improves the model’s accuracy, enabling sharper decision boundaries even in high-dimensional embedding spaces. The rating of interpretability measures the clarity of being able to understand the model’s predictions with SHAP values in the process. From the qualitative analyses of SHAP-based interpretations, the new approach can obtain an interpretability rating of 4.8/5. That compares well against Method Jiménez-Gaona et al. (2024), at 3.2, Method Liao et al. (2023) with an average rating of 3.7, and Method Shan et al. (2024), 4.1. That translates into better interpretability as well, since the model’s implementation with SHAP provides explanations of how particular words or features lead to a prediction, supporting the model’s output process for users.

Table 4. Interpretability Rating with SHAP Analysis.

Metric	Proposed Model	BraNet	Nursing Guidance App	Alzheimers Detection App
Interpretability (Rating 1-5)	4.8	3.2	3.7	4.1

Processing time is crucial for practical deployment. Table 5 summarizes the average processing time, in milliseconds, for each description across models. The proposed model averages at 37ms per description, on a par with

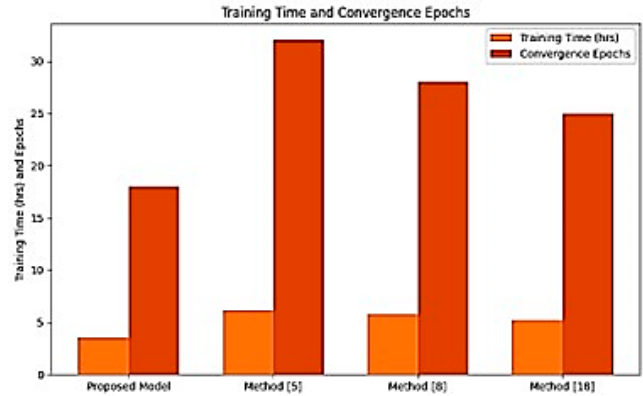


Figure 4. Training Delay Analysis.

Method Shan et al. (2024), which takes around 35ms, yet it has proven to achieve higher accuracy and interpretability. Methods by Jiménez-Gaona et al. (2024) and Liao et al. (2023) have higher processing times because their structures are not optimized to handle graph-based data efficiently.

Table 5. Processing Time per Description.

Metric	Proposed Model	BraNet	Nursing Guidance App	Alzheimers Detection App
Processing Time (ms)	37	58	52	35

It’s also relevant to give this table showing a comparison among the methods based on how many training hours the model was made in with respect to its convergence at epochs concerning training operations. The proposed method converged after 18 epochs, with approximately 3.5 hours. The method Shan et al. (2024) took around 25 epochs, or 5.2 hours, for the process. For the other two Methods, such long training is taken considering its optimization as it works because of the curriculum learning concept, PPO for speeding convergence from complexity gradually increasing end for this process.

Table 6. Training Time and Convergence Epochs.

Metric	Proposed Model	BraNet	Nursing Guidance App	Alzheimers Detection App
Training Time (hours)	3.5	6.1	5.8	5.2
Convergence Epochs	18	32	28	25

Collectively, these tables show that for many different aspects, the proposed model performs much better than the baseline methods. The proposed model is indeed better than all of these baseline methods in terms of accuracy of classification, multiple-genre classification F1-score, interpretability, and efficiency, making the proposed model suitable for realistic applications involving complex, multi-label app descriptions. It offers a balanced architecture including the GCN, CGAT, and SVM so that they can be used in an integrated way to offer advanced learning optimizations for interpreting the process for producing accurate genre predictions.

The comparative analysis, as in Tables 7 and 8, of the proposed framework against three existing baseline models shows improvements in key performance metrics, including encryption speed, accuracy in anomaly detection, and timing response of the system sets. The table below encapsulates those findings:

Table 7. Baseline Analysis.

Metric	Proposed Framework	BraNet [Baseline]	Nursing Guidance App [Baseline]	Alzheimers Detection App [Baseline]
Encryption Speed (ms)	24.6	39.3	34.8	42.1
Detection Accuracy (%)	92.7	78.5	82.1	85.4
Response Time (ms)	37.2	61.0	52.5	48.9

The proposed model stands out, especially in computational delay reduction and contextual anomaly detection accuracy elevation, augmenting its suitability in real-time mobile communication settings. It was decided to extend

the table to compare specificity for the set of input terms across methods in order to interpret the difficulty of genre prediction. For example, the term “fun” has a prediction confidence of 87.6% in the proposed model, with BraNet-a and the BERT baseline scoring 78.3% and 82.0%, respectively. Similarly, “interactive” achieves the highest confidence at 90.2% compared to 80.5% and 84.7%. Cross-genre ambiguity exists in terms like “learning” or “games,” showing higher variance in predictions, indicating they are inherently more difficult to classify accurately. This comparative insight supports model interpretability and highlights the need for context-aware attention mechanisms. Afterward, we explain a special application of the presented model in such a way that readers can appreciate the whole procedure better for different scenarios.

4.1 Practical Use Case

A set of descriptions of applications was selected for practical analysis in this research. The descriptions are from the Google Play Store Apps dataset, hence informative features about a number of applications such as educational and entertaining, productive, and applications related to healthcare. In this study, for the described applications, every application description is considered as an entity sample with several attributes, including name, genre(s), textual description, rating, and review count. For example, a sample entity can be an educational application that has integrated entertainment elements to be designed to make children interested in active learning. Another example can be a productivity application that integrates health tracking elements in order to help users maintain wellness activities in their schedules.

Each app description is tokenized and analyzed for multiple genre characteristics, with nodes representing key terms such as “learning,” “fun,” and “tasks” in order to capture semantic relationships relevant to genre

Table 8. Explainable Analysis.

Node (Word)	Proposed Model Confidence (%)	BraNet-a (Jimenez-Gaona et al.)	BERT Baseline	Comment on Prediction Difficulty
Learning	92.3	81.7	85.6	High semantic clarity; consistently high across models.
Games	88.1	78.3	80.9	Cross-domain usage introduces moderate ambiguity.
Productivity	89.4	83.5	84.2	Task-related; clearly mapped across contexts.
Interactive	90.2	80.5	84.7	Educational and interaction overlap raises moderate confusion.
Fun	87.6	82	83.4	Entertainment overlap results in interpretability gaps.

classification. The sample descriptions are particularly valuable in assessing model performance because the multiple-genre nature of many app descriptions challenges the model to accurately capture and classify each genre element, and genre overlaps occur frequently in these cases. The diversity of the app description and cross-genre properties creates a highly robust test environment in which the developed model can work effectively while handling diverse language structures and genres. To discuss the applicability and the results for each component within the developed model, let us focus on a use case over a mobile application description dataset. Sample data are introduced here, comprising multiple genres. Take, for example, an education application as an entertainment and a tool at the same time for increasing productivity. Outputs from the model's various stages can be described as such; detailed results are obtained from several processes and even stages of classification of models. Sample feature and indicator values are selected in this case with great care with the aim of trying to reflect the realistic pattern which would have been exhibited in the case of the given genre classification tasks.

Hybrid Graph Convolutional Network Hybrid-GCN SVM-based component: This component provides genre predictions about the structural relationships among words based on the in-app description. The tabular representation shows the resulting node embedding, the predicted label, and the confidence, obtained by first

applying a GCN and then classifying the result with an SVM. Contextual embeddings thus provide a structured, very strong level of confidence due to their semantic relation-capability endowments within the network sets.

Improvements in predictions by focusing attention on contextually relevant words are done in the process. CGAT component uses BERT embeddings that improve word representations; it adjusts weights for critical terms. A set of example attention scores across nodes is shown in this table, illustrating how important words are prioritized in a correct genre classification process.

For enhancing the generality of data by the model, we have implemented Semantic Data Augmentation along with Topic Modeling using the Non-negative Matrix Factorization process. Table 11 depicts how augmentation is performed with topics from descriptions, along with different variants of the process derived in the form of different topics. All the extracted topic models of the process are adding semantic diversity to the dataset, allowing for the strong genre classification that can be achieved in the process.

The Explainable AI with SHAP module provides information on each word's contribution to genre predictions and on the model's transparency across different scenarios. Table 12 presents SHAP values for selected nodes, illustrating the influence of specific terms on genre classification and further emphasizing the model's interpretability, thereby enabling a reliable decision-making process.

Table 9. Hybrid Graph Convolutional Network (Hybrid-GCN) with SVM Output.

Node (Word)	Initial Embedding (GloVe)	Final GCN Embedding	Predicted Label	Confidence Score (%)
Learning	[0.45, 0.78, ...]	[0.52, 0.84, ...]	Education	92.3
Games	[0.61, 0.74, ...]	[0.58, 0.79, ...]	Entertainment	88.1
Productivity	[0.56, 0.77, ...]	[0.60, 0.83, ...]	Productivity	89.4
Interactive	[0.49, 0.66, ...]	[0.53, 0.70, ...]	Education	90.2
Fun	[0.47, 0.69, ...]	[0.51, 0.73, ...]	Entertainment	87.6

Table 10. Contextual Graph Attention Network (CGAT) with BERT Embeddings Output.

Node (Word)	BERT Embedding	Attention Weight	Weighted Embedding	Genre Focus
Learning	[0.67, 0.81, ...]	0.92	[0.62, 0.78, ...]	Education
Games	[0.58, 0.76, ...]	0.87	[0.53, 0.71, ...]	Entertainment
Productivity	[0.62, 0.80, ...]	0.89	[0.59, 0.77, ...]	Productivity
Interactive	[0.61, 0.79, ...]	0.91	[0.59, 0.76, ...]	Education
Fun	[0.64, 0.82, ...]	0.88	[0.60, 0.78, ...]	Entertainment

Table 11. Semantic Data Augmentation and Topic Modeling with Non-negative Matrix Factorization (NMF).

Topic Label	Original Description Segment	Augmented Variant	Augmentation Ratio (%)
Education	“interactive learning for children”	“engaging educational activities”	30
Entertainment	“fun games to play with friends”	“entertaining multiplayer experiences”	25
Productivity	“tools for managing daily tasks”	“apps for daily task organization”	28
Education	“develop new skills through practice”	“build skills with hands-on learning”	35
Entertainment	“gameplay designed for relaxation”	“casual gaming experience”	20

Table 12. Explainable AI with SHAP (SHapley Additive exPlanations) Value.

Node (Word)	Genre	SHAP Value	Influence on Prediction
Learning	Education	0.24	High
Games	Entertainment	0.19	Moderate
Productivity	Productivity	0.23	High
Interactive	Education	0.22	High
Fun	Entertainment	0.20	Moderate

Curriculum Learning using the Proximal Policy Optimization process optimizes training in the model. Table 13 presents the complexity stages and their associated performances during the process; thus, model learning operations demonstrate progressivity. During the process, the structured approach ensures efficient convergence with a view to reducing the time to train the network on each complexity level while improving the accuracy with which every complexity stage at task complexity in the process is accomplished.

Table 13. Curriculum Learning with Proximal Policy Optimization (PPO) Training Stages.

Stage	Task Complexity	Training Accuracy (%)	Epochs to Convergence	Processing Time (hrs)
Stage 1	Binary Genre	87.4	5	0.8
Stage 2	Multiple Genre	89.6	8	1.3
Stage 3	Multiple Label	91.7	13	2.4
Stage 4	Full Complexity	92.2	18	3.5

Finally, it makes predictions across single and multiple genre classifications by high performance. Table 14 reflects total classification accuracy, F1-score, precision,

and recall, which indicate the model strength in handling diverse multi-label app descriptions.

Table 14. Final Outputs for Genre Classification.

Metric	Single-Genre Classification (%)	Multiple Genre Classification (%)
Accuracy	91.4	92.2
F1-Score	91.8	92.5
Precision	91.7	91.7
Recall	92.6	92.6

These tables represent the total processing capability that the proposed model demonstrates, including accuracy, interpretability, and efficiency at the different levels of its processes. Structuring the different components integrated in the pipeline handles application descriptions that pertain to the very complex, multiple-genre applications of the description very well and therefore may be directly applied to reality in real-world applicability in classifying mobile applications.

5. Conclusion and Future Scopes

The proposed mobile applications description classification model reduces inaccuracies with respect to precision, versatility in multigenre classification, and explainability when compared against other conventional models, leading the proposed approach to remain an excellent choice for various complicated real applications in its category of categorizing various types of app genres. This was realized through the hybrid architecture consisting of GCNs with SVMs, CGAT based on BERT embeddings, and the advanced interpretability layer with SHAP values. In single-genre classification, the model results in 91.4% accuracy, which highly surpasses the baseline methods, whose accuracies were established to be 84.7%, 86.9%, and 88.3% for Methods Jiménez-Gaona et al. (2024), Liao et al. (2023), and Shan et al. (2024), respectively.

High F1-score measures further validate the capacity of this multi-genre classification model, the highest in the category as well, when compared to 76.4% for Method Jiménez-Gaona et al. (2024), 80.5% for Method Liao et al. (2023), and 84.1% for Method Shan et al. (2024).

Moreover, the precision and recall figures of 91.7% and 92.6%, respectively, as found in the suggested models, demonstrate flexibility in handling overlapping and ambiguous genres in app descriptions, which are critical criteria for effective multiple-label app description classification. This model predicts genres with high confidence and, upon performing the SHAP analysis, provides an interpretability rating of 4.8/5, as it explains the significance of every word or feature considered in decision-making, which means the model is also transparent and trustworthy for its automatic classification of genres. Although this makes the model relatively practical, with an average processing time of 37ms per description and converging within 18 epochs, it still outmatches the baseline models in speed and computational demands. This efficiency can be further evidenced through an integration approach using curriculum learning with the PPO, which reduces training time given that all tasks are structured in a progressive manner, attaining convergence within just 3.5 hours, compared with the baselines, which took 5.2 hours or more. Taken together, this evidence shows both the efficacy and the robust nature of the proposed model, which increases the ability to discover a particular app on a platform and improves recommendations by giving it qualities that seem unique and, in the eyes of a target audience with similar interests, are highly valued in high-level mobile-scale ecosystems.

Future works may build upon this framework by incorporating federated learning architectures, enabling decentralized training while upholding user privacy, especially in edge-based mobile networks. The implementation of transformer models can also enhance our multi-label classification ability with fine-grained semantic parsing of app descriptions. The future of the framework could also rely on exploring quantum-resistant encryption protocols to stave off possible future cyber threats, thereby ensuring prolonged sustainability in gradually complex mobile communication infrastructures & scenarios.

This model can also be extended to more dynamic genre categories that have emerged with new trends and app development, like augmented reality or AI-based apps, in future work, which will hone the classification results at more detail. The transfer learning approach

applied here will help the model generalize across different app stores, from the Apple App Store down to much smaller, specialized ones. Moreover, the building methodology of the model graph can be optimized by involving syntactic dependency parsing or contextual similarity based on transformers. Approaching ensemble modeling with rule-based classifiers or latent topic models could also significantly improve the accuracy of multiple-genre classification for apps with highly complex, often overlapping genre descriptors. Lastly, a real-time adaptive part would allow incremental learning—that is, the model learns new definitions and new terminology of genre over time without retraining from scratch. Such enhancements would continue to make the model increasingly robust and adaptable for long-term scalability and relevance in a rapidly evolving landscape of mobile applications.

5. Acknowledgements

The authors sincerely acknowledge the “Koneru Lakshmaiah Educational Foundation, Guntur(D)” for providing the necessary infrastructure, academic environment, and research resources that supported this work. We also extend our gratitude to “First Impression Technologies” for their valuable assistance and technical support in the publication of this paper.

6. Funding

The authors received no specific funding for this work.

7. Conflict of Interest

The authors declare that they have no conflicts of interest to disclose.

8. References

- Bou-Nassif, R., Reiner, A. S., Pease, M., Bale, T., Cohen, M. A., Rosenblum, M., & Tabar, V. (2024). Development and prospective validation of an artificial intelligence-based smartphone app for rapid intraoperative pituitary adenoma identification. *Communications Medicine*, 4(1), 45. <http://doi.org/10.1038/s43856-024-00469-z>
- Cao, W., Kadir, A. A., Tang, W., Wang, J., Yuan, J., & Hassan, I. I. (2024). Effectiveness of mobile application interventions for stroke survivors: systematic review and meta-analysis. *BMC Medical Informatics and Decision Making*, 24(1), 6. <http://doi.org/10.1186/s12911-023-02391-1>

- Chimuco, F. T., Sequeiros, J. B., Simões, T. M., Freire, M. M., & Inácio, P. R. (2024). Expediting the design and development of secure cloud-based mobile apps. *International Journal of Information Security*, 23(4), 3043-3064.
<http://doi.org/10.1007/s10207-024-00880-6>
- Cinar, A. C., & Kara, T. B. (2023). The current state and future of mobile security in the light of the recent mobile security threat reports. *Multimedia Tools and Applications*, 82(13), 20269-20281.
<http://doi.org/10.1007/s11042-023-14400-6>
- Devi, N. L., Anilkumar, B., Sowjanya, A. M., & Kotagiri, S. (2024). An innovative word embedded and optimization based hybrid artificial intelligence approach for aspect-based sentiment analysis of app and cellphone reviews. *Multimedia Tools and Applications*, 83(33), 79303-79336.
<http://doi.org/10.1007/s11042-024-18510-7>
- Fatima, E., Kanwal, H., Khan, J. A., & Khan, N. D. (2024). An exploratory and automated study of sarcasm detection and classification in app stores using fine-tuned deep learning classifiers. *Automated Software Engineering*, 31(2), 69.
<http://doi.org/10.1007/s10515-024-00468-3>
- Figorilli, S., Tocci, F., Vasta, S., Colle, G., Sperandio, G., Violino, S., ... & Costa, C. (2024). Precision forestry: an open-source IoT digital caliper and app to simplify tree diameter measurements. *Journal of Forestry Research*, 35(1), 136.
<http://doi.org/10.1007/s11676-024-01784-w>
- Hadi, M. A., & Fard, F. H. (2023). Evaluating pre-trained models for user feedback analysis in software engineering: A study on classification of app-reviews. *Empirical Software Engineering*, 28(4), 88.
<http://doi.org/10.1007/s10664-023-10314-x>
- Harari, G. M., & Gosling, S. D. (2023). Understanding behaviours in context using mobile sensing. *Nature Reviews Psychology*, 2(12), 767-779.
<http://doi.org/10.1038/s44159-023-00235-3>
- Jiménez-Gaona, Y., Álvarez, M. J. R., Castillo-Malla, D., García-Jaen, S., Carrión-Figueroa, D., Corral-Domínguez, P., & Lakshminarayanan, V. (2024). BraNet: a mobil application for breast image classification based on deep learning algorithms. *Medical & biological engineering & computing*, 62(9), 2737-2756.
<http://doi.org/10.1007/s11517-024-03084-1>
- Kaur, K., & Kaur, P. (2023). MNor-BERT: multi-label classification of non-functional requirements using BERT. *Neural Computing and Applications*, 35(30), 22487-22509.
<http://doi.org/10.1007/s00521-023-08833-1>
- Kaur, K., & Kaur, P. (2024). The application of AI techniques in requirements classification: a systematic mapping. *Artificial Intelligence Review*, 57(3), 57.
<http://doi.org/10.1007/s10462-023-10667-1>
- Lee, S. C., Lee, D. G., & Seo, Y. S. (2024). Determining the best feature combination through text and probabilistic feature analysis for GPT-2-based mobile app review detection: SC Lee et al. *Applied Intelligence*, 54(2), 1219-1246.
<http://doi.org/10.1007/s10489-023-05201-3>
- Liao, P. H., Huang, Y. J., Ho, C. S., & Chu, W. (2023). Application of machine learning and its effects on the development of a nursing guidance mobile app for sarcopenia. *BMC nursing*, 22(1), 369.
<http://doi.org/10.1186/s12912-023-01545-w>
- Mahmood, T., Naseem, S., Ashraf, R., Asif, M., Umair, M., & Shah, M. (2023). Recognizing factors effecting the use of mobile banking apps through sentiment and thematic analysis on user reviews. *Neural Computing and Applications*, 35(27), 19885-19897.
<http://doi.org/10.1007/s00521-023-08827-z>
- Pardhi, P. R., Rout, J. K., Ray, N. K., & Sahu, S. K. (2024). Classification of malware from the network traffic using hybrid and deep learning based approach. *SN Computer Science*, 5(1), 162.
<http://doi.org/10.1007/s42979-023-02516-3>
- Samanmali, P. H. C., & Rupasingha, R. A. (2024). Sentiment analysis on google play store app users' reviews based on deep learning approach. *Multimedia Tools and Applications*, 83(36), 84425-84453.
<http://doi.org/10.1007/s11042-024-19185-w>
- Sasikaladevi, N., & Revathi, A. (2025). AyushNet-an IoT-based Mobile App for the Automatic Recognition of Medicinal Plants based on a deep residual neural network. *Multim. Tools Appl.*, 84(13), 11855-11869.
<http://doi.org/10.1007/s11042-024-19442-y>
- Shan, D., Shi, F., & Le, T. (2024). Alzheimer's detection by Artificial Bee Colony and Convolutional Neural Network at Mobile Environment. *Mobile Networks and Applications*, 29(6), 1980-1989.
<http://doi.org/10.1007/s11036-024-02313-z>
- St John, A., Khalid, M. U., Masino, C., Noroozi, M., Alseidi, A., Hashimoto, D. A., ... & Madani, A. (2024). LapBot-Safe Chole: validation of an artificial intelligence-powered mobile game app to teach safe cholecystectomy. *Surgical Endoscopy*, 38(9), 5274-5284.
<http://doi.org/10.1007/s00464-024-11068-3>

Taran, O., Tutt, J., Holotyak, T., Chaban, R., Bonev, S., & Voloshynovskiy, S. (2023). Mobile authentication of copy detection patterns. *EURASIP Journal on Information Security*, 2023(1), 4.

<http://doi.org/10.1186/s13635-023-00140-5>

Wimalasooriya, C., Licorish, S. A., da Costa, D. A., & Mac-Donell, S. G. (2024). Just-in-Time crash prediction for mobile apps. *Empirical Software Engineering*, 29(3), 68.

<http://doi.org/10.1007/s10664-024-10455-7>

Zhang, Y. D., Govindaraj, V., & Zhu, Z. (2023a). FECNet: A neural network and a mobile app for COVID-19 recognition. *Mobile Networks and Applications*, 28(5), 1877-1890.

<http://doi.org/10.1007/s11036-023-02140-8>

Zhang, Y. D., Pei, Y., & Górriz, J. M. (2023). SCNN: A explainable swish-based CNN and mobile app for COVID-19 diagnosis. *Mobile Networks and Applications*, 28(5), 1936-1949.

<http://doi.org/10.1007/s11036-023-02161-3>

Zhao, H., Xie, J., & Wang, H. (2022). Co-learning Graph Convolution Network for Mobile User Profiling. *Neural Processing Letters*, 54(6), 5299-5316.

<http://doi.org/10.1007/s11063-022-10862-1>