PROCEEDINGS OF
INTERNATIONAL CONFERENCE ON INTELLIGENT SYSTEMS AND NEW APPLICATIONS
https://proceedings.icisna.org/
3rd International Conference on Intelligent Systems and New Applications (ICISNA'25), Antalya, December 12-14, 2025.

Deep Learning Methods for The Classification of Turkish Music Genres

Muhammed EMINOGLU[1], Murat KOKLU[2]

[1]*Institute of Natural Sciences, Department of Computer Engineering, Selcuk University, Türkiye*
muhammed.eminoglu@selcuk.edu.tr, ORCID: 0009-0000-8387-6590

[2] *Technology Faculty, Department of Computer Engineering, Selcuk University, Türkiye*
mkoklu@selcuk.edu.tr, ORCID: 0000-0002-2737-2360

***Abstract*— Accurate classification of music genres is essential for the effective management of digital music archives and for improving the reliability of recommendation systems. Traditional approaches based on audio signal analysis often fail to utilize the rich semantic and structural information embedded in song lyrics. In this study, a deep learning–based method is proposed for the automatic identification of music genres by using Turkish song lyrics. An original dataset consisting of four thousand songs collected from real-world sources and balanced to eliminate class imbalance was constructed. Comprehensive normalization procedures compatible with Turkish morphology were applied during the text preprocessing stage. The classification performance was evaluated using Convolutional Neural Networks, Long Short-Term Memory networks, Transformer-based architectures, and a pretrained Turkish Contextual Language Representation model. Additionally, to assess the performance of these models relative to large-scale language models, the Llama-3-70B model was tested using a direct inference approach without any additional training. Furthermore, a weighted ensemble learning architecture that integrates the predictions of different models was developed. Experimental results show that among the individual models, the Turkish Contextual Language Representation model achieved the highest accuracy. However, the proposed ensemble learning architecture outperformed all single deep learning models and the Llama-3-70B model, achieving 68.17 percent accuracy, 0.68 F1-score, 0.69 precision, and 0.67 recall. Genre-specific results indicate that the Rap genre exhibited the highest discriminability with an F1-score of 0.92, whereas Pop (0.61 F1), Rock (0.58 F1), and Arabesque (0.57 F1) displayed notable overlaps in lyrical and thematic characteristics.**

***Keywords*— Music Genre Classification; Natural Language Processing; Deep Learning; Ensemble Learning; TurkBERT; Llama-3; Text Mining.**

I. INTRODUCTION

With the proliferation of digital music platforms, it has become essential to effectively categorize, index, and recommend millions of songs to users [1]. Music genre classification is one of the cornerstones of this process and directly impacts the performance of information retrieval systems [2]. While traditional approaches focus on acoustic features extracted from music files, such as Mel-frequency cepstral coefficients (MFCC), spectral center, and rhythm [3], lyrics provide complementary and distinctive information about the emotional mode, thematic content, and narrative structure of a piece [4].

Today, artificial intelligence and machine learning techniques offer high success rates in analyzing complex data sets and solving classification problems. These methods are effectively used in many different disciplines, ranging from image processing and signal analysis to the classification of agricultural products [28, 30, 34], industrial engineering problems [29], food safety [31], and medical diagnosis systems [32-33]. The methodological successes achieved in these studies also shed light on areas such as natural language processing and music genre classification.

Song lyrics constitute a unique subfield of natural language processing (NLP). Mayer and colleagues [5] emphasized that song lyrics require approaches beyond standard text mining methods due to their poetic structure, metaphorical expression, and repetitive nature. This analysis becomes even more complex in languages with rich morphological structures and agglutinative characteristics, such as Turkish [6]. However, studies conducted on different datasets in the literature show that artificial intelligence and machine learning-based classification methods achieve high success rates and produce effective results [26]. While classification studies based on song lyrics have been concentrated in the literature [7, 8], deep learning-based comparative analyses on Turkish music data sets (corpora) are quite limited.

Song lyrics constitute a unique subfield of natural language processing (NLP). Mayer and colleagues [5] emphasized that song lyrics require approaches beyond standard text mining methods due to their poetic structure, metaphorical expression, and repetitive nature. This analysis becomes even more complex in languages with rich morphological structures and agglutinative characteristics, such as Turkish [6]. However, studies conducted on different datasets in the literature show that artificial intelligence and machine learning-based classification methods achieve high success rates and produce effective results [26]. While classification studies based on song lyrics have been concentrated in the literature [7, 8], deep learning-based comparative analyses on Turkish music data sets (corpora) are quite limited.

II. LİTERATURE REVİEW

Music genre classification has been an active area of research since the early 2000s. Tzanetakis and Cook [14] achieved 61% accuracy using acoustic features such as Mel-frequency cepstral coefficients (MFCC), spectral center, and zero-crossing rate on the GTZAN dataset, and this work has become a reference point in the field. Li et al. [15] combined Daubechies wavelet coefficients and MFCC features with Support Vector Machine (SVM) and reported a classification success rate of 78.5% across 10 music genres. While these studies demonstrated the potential of acoustic features in genre classification, they disregarded the semantic information carried by song lyrics.

In text-based approaches, a different perspective has been adopted. Fell and Sporleder [7] conducted experiments using Naive Bayes and SVM on a 78,000-song English dataset, employing bag-of-words, n-grams, and stylistic features (average word length, repetition rate, rhyme structure); they determined that stylistic features alone performed similarly to content features. Tsaptsinos [8] modeled song lyrics at both the word and line levels using hierarchical attention networks, achieving results that outperformed traditional CNN and LSTM architectures. Oramas et al. [17] developed a multimodal CNN architecture combining audio spectrograms, lyrics, and album covers on the MuMu dataset; they showed that this hybrid approach achieved a 6-8% higher F1-score compared to single modalities. Malheiro et al. [16] classified four emotional categories with 64% accuracy on 180 songs using sentiment features (valence, arousal, tension) extracted from lyrics, proving the contribution of lyrical sentiment analysis to genre detection.

Research in the field of Turkish NLP has remained relatively limited. Özkan and Kar [18] performed multi-class classification by applying the BERT deep learning technique on academic and scientific texts written in Turkish over the past 10 years; they reported that the resulting system achieved an accuracy rate of 96%. Schweter [19] presented the BERTurk model, trained on a 35GB Turkish text corpus, achieving significant improvements over multilingual BERT models in sentiment analysis and text classification tasks.

In terms of large-scale datasets and different approaches, Defferrard et al. [22] introduced the Free Music Archive (FMA) dataset, consisting of 106,574 tracks, providing a comprehensive resource for music information retrieval research. Ferraro and Lemström [23] performed large-scale genre classification through the automatic identification of recurring patterns in symbolically encoded music. Specifically for Turkish music classification, Hızlısoy and Tüfekci [24] achieved 91.72% accuracy on a unique dataset of 200 Turkish songs using the Convolutional Long Short-Term Memory Deep Neural Network (CLDNN) architecture. Durdağ and Erdoğmuş [25] converted music files into Mel-spectrogram images and classified them using deep learning networks, demonstrating the effectiveness of visual representations in sound-based classification.

A review of the literature reveals that there is no comprehensive comparative analysis based on deep learning for Turkish song lyrics. Existing studies mostly focus on English datasets, ignoring the rich morphological structure and agglutinative characteristics of Turkish. Furthermore, the performance of large language models (LLMs) in lyric-based classification has not yet been sufficiently researched. This study aims to fill these gaps.

III. MATERIALS AND METHODS

This section details the methodological framework of the hybrid artificial intelligence system developed for music genre identification from Turkish song lyrics. The study follows a systematic flow consisting of four main stages, from data preparation to model deployment: (1) Data Collection and Balancing, (2) NLP-Based Preprocessing, (3) Deep Feature Extraction and Classification, and (4) Decision Combining with Ensemble Learning.

A. Dataset

The dataset used in this study was compiled using web scraping methods from popular online music platforms in Turkey. The raw dataset initially exhibited significant class imbalance. To prevent model bias, the number of data points for each music genre (Pop, Rock, Rap, Arabesk) was fixed at 1,000, creating a balanced dataset of 4,000 songs in total. The dataset was split into training (70%), validation (15%), and test (15%) sets. The distribution of songs in the dataset by music genre is shown in Figure 1.

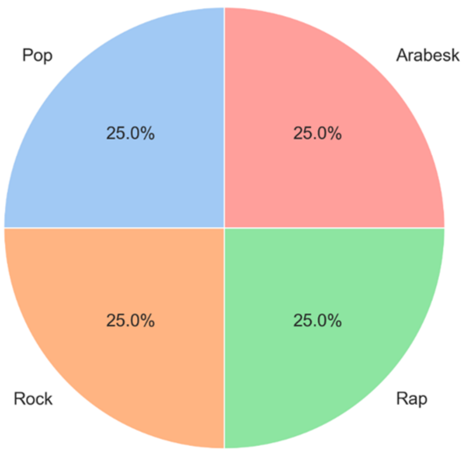


Fig. 1 Distribution of songs in the dataset by music genre (4 Classes x 1000 Songs)

B. Section Headings

No more than 3 levels of headings should be used. All headings must be in 10pt fonts. As with the title, every wocrd in headings should be capitalized except for minor words.

1) Text Cleaning and Normalization: During the data preprocessing stage, all characters in the text were converted to lowercase (case folding) and punctuation marks such as commas, periods, and exclamation points were removed. Additionally, numerical expressions and special characters (@, #, &, etc.) were cleaned, and parenthetical expressions frequently found in song lyrics were also adjusted. Finally, repeated spaces and line break characters were normalized to prepare the text for analysis.

2) Turkish-Specific Operations: Veri setindeki Türkçe karakterler (ç, ğ, ı, ö, ş, ü) korunmuş ve herhangi bir ASCII dönüşümü uygulanmamıştır. Süreç kapsamında Türkçe etkisiz kelimeler (stop-words) isteğe bağlı olarak filtrelenirken, metnin özgün yapısını yansıtmak amacıyla argo ve günlük konuşma dili ifadeleri ise muhafaza edilmiştir.

3) *Tokenization:* A 30,000-word vocabulary was created by applying word-level tokenization for classical models such as CNN, LSTM, and Transformer, and expressions not included in the vocabulary were represented by the <UNK> token. In the TurkBERT model, WordPiece tokenization was used to minimize the out-of-vocabulary (OOV) problem by splitting unknown words into subword units.

4) Sequence Length and Padding: The maximum sequence length at model inputs is set to 256 tokens. To ensure standard length, zero-padding is applied to short texts, while long texts exceeding the limit are subject to truncation. Looking at the dataset statistics, the average song lyric length is calculated to be 187 words.

C. Model Architectures

The study compares five different approaches that address the problem from different angles. Deep learning architectures demonstrate superior performance compared to traditional methods, particularly in feature extraction from large datasets. Studies have reported that models based on Convolutional Neural Networks (CNN), in particular, have achieved accuracy rates of up to 98% in the classification of visual and structural data [35]. This study also tested the success of similar deep architectures on text-based data. Schindler et al. [20] compared shallow and deep neural network architectures in music genre classification and demonstrated the superiority of deep models. In light of these findings, the following architectures were selected.

1) *Convolutional Neural Networks (CNN):* Convolutional Neural Networks (CNNs) are designed to capture local n-gram patterns within text [9]. From image processing to signal analysis, CNN architectures are known to deliver superior performance in feature extraction and classification tasks, while deep learning-based approaches successfully model distinctive features in complex data structures [27]. The model architecture consists of the following components:

- Embedding Layer: 128-dimensional word vectors
- 1D Convolution Layers: 128 filters with 3 parallel filter sizes (3, 4, 5 kernels)
- Max-Pooling: Selection of the most prominent features from each filter output
- Dropout: Prevention of overfitting at a rate of 0.5
- Dense Layer: 4-class softmax output

This architecture learns language usage specific to the genre by identifying local word patterns in expressions such as “when I look into your eyes.”

2) BiLSTM (Bidirectional LSTM): Bidirectional Long Short-Term Memory networks model contextual relationships by processing text in both forward and backward directions [10]. Model structure:

- Embedding Layer: 128-dimensional word vectors
- Bidirectional LSTM: 64-unit bidirectional LSTM layer (total of 128 outputs)
- Attention Mechanism: Focusing on important words
- Dense Layers: 64-unit hidden layer + 4-class output

BiLSTM has the capacity to capture long-term dependencies in song lyrics (e.g., thematic consistency between the chorus and the verse).

3) Transformer: The Transformer architecture analyzes relationships between words independently of distance using a self-attention mechanism [11]. Applied structure:

- Positional Encoding: Adding sequence position information
- Multi-Head Attention: 8-head attention mechanism
- Feed-Forward Network: 256-unit feed-forward network
- Layer Normalization: Layer normalization
- Encoder Blocks: 2-layer encoder structure

This architecture simultaneously evaluates the semantic relationships between all word pairs in the song lyrics.

4) *TurkBERT (Transfer Learning):* The pre-trained BERT model for Turkish (BERTurk) has been adapted using a transfer learning approach [19]. Fine-tuning process:

- Base Model: dbmdz/bert-base-turkish-cased (110M parameters)
- Maximum Sequence Length: 256 tokens
- Learning Rate: 2e-5 (AdamW optimizer)
- Batch Size: 16
- Number of Epochs: 5 (with early stopping)
- Classification Header: Dense layer added to the [CLS] token output

TurkBERT's pre-trained Turkish language knowledge delivers high performance even with limited data.

5) *Llama-3-70B (Zero-Shot Comparison):* The Llama-3-70B large language model developed by Meta was tested using the “zero-shot” method without seeing any training data. The prompt structure used:

```
json
{
  “task”: “music_genre_classification”,
  “language”: “Turkish”,
  “instruction”: “Determine the music genre of the following Turkish song lyrics.”,
  “options”: [“Pop”, “Rock”, ‘Rap’, “Arabesk”],
  “lyrics”: “{lyrics}”,
  “output_format”: “Return only the genre name.”
}
```

This comparison aims to compare the performance of general-purpose large language models on domain-specific tasks with specialized models.

D. *Ensemble Learning Strategy*

The advantages of community systems in decision-making processes have been examined in detail in the literature [21]. In this study, the Weighted Soft Voting strategy was applied to minimize the errors of individual models and combine the strengths of different architectures [13].

1) *Ensemble Structure:* Four basic models (CNN, BiLSTM, Transformer, TurkBERT) have been included in the ensemble system. The weight coefficient has been determined according to the performance of each model in the validation set. The coefficients are given in Table 1.

TABLE I
WEIGHTING COEFFICIENTS DETERMINED FOR THE ENSEMBLE STRUCTURE MODEL

Model	*Validasyon Doğruluğu*	*Ağırlık (w)*
CNN	*0.57*	*0.20*
BiLSTM	*0.51*	*0.15*
Transformer	*0.57*	*0.20*
TurkBERT	*0.67*	*0.45*
TOPLAM	*-*	*1.00*

2) *Decision Mechanism:* The final prediction was calculated as the weighted sum of each model's softmax probability outputs:

$$P_{ensemble}(y=c)=\sum_{i=1}^{4} w_i \cdot P_i(y=c)$$

Here, $P_i(y=c)$ represents the probability value produced by the ith model for class c; w_i represents the weight coefficient of the corresponding model.

E. *Performance Evaluation Metrics*

The following metrics were used to objectively compare model performance:

1) *Accuracy:* Represents the ratio of correct predictions among all predictions:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

2) *Precision:* Measures how many positive predictions are actually positive:

Precision = TP / (TP + FP)

3) *Recall:* Shows how many true positives were correctly detected:

Recall = TP/(TP+FN)

4) *F1-Score:* The harmonic mean of Precision and Recall, providing a more reliable metric for imbalanced classes:

   F1 = 2×(Precision×Recall)/(Precision+Recall)

5) Confusion Matrix: A matrix showing the distribution of actual and predicted labels for each class. Rows represent actual classes, while columns represent predicted classes. Diagonal elements indicate correct classifications, while off-diagonal elements indicate errors.

IV. Findings and Analysis

This section presents the results of experimental studies conducted to evaluate the performance of the proposed music genre classification system. First, the hardware and software infrastructure used to conduct the experiments is detailed, followed by a comparative analysis of the success rates of the developed deep learning models (CNN, LSTM, Transformer, TurkBERT) and the Llama-3 model. Finally, the class-based performance of the proposed Ensemble model and the cross-type confusion matrix are examined.

A. *Model Performance Comparison*

The training and testing processes for all models were conducted on an NVIDIA CUDA-supported workstation to meet high-performance computing requirements. Python 3.10 programming language and PyTorch 2.1 deep learning library were used as the software infrastructure. The Hugging Face Transformers library was utilized for the TurkBERT and Llama-3 models. Details of the hardware and software components used in the experimental studies are presented in Table 2. Thanks to this powerful hardware infrastructure, the fine-tuning of the TurkBERT model and the optimization of the Ensemble model were completed efficiently.

TABLE 2
HARDWARE AND SOFTWARE FEATURES USED IN EXPERIMENTAL STUDIES

Bileşen	*Özellikler*
GPU	*NVIDIA GeForce RTX 4060 (8GB GDDR6, CUDA 12.1)*
CPU	*Intel Core i7-13700H (14 Çekirdek, 5.0 GHz)*
RAM	*32 GB DDR5 4800 MHz*
İşletim Sistemi	*Windows 11 Pro*
Yazılım	*Python 3.10, PyTorch 2.1, Transformers 4.35*

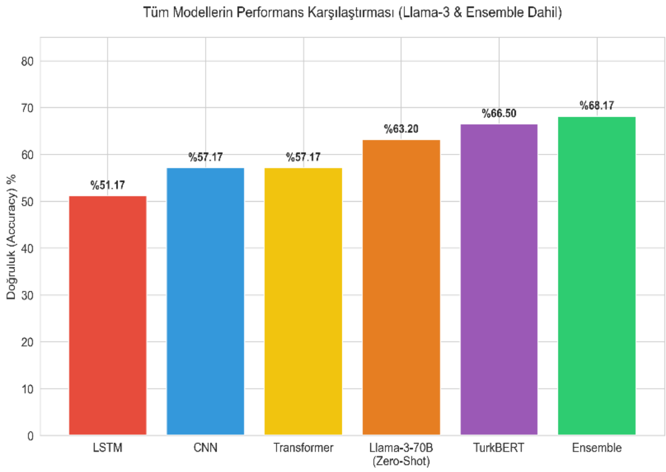


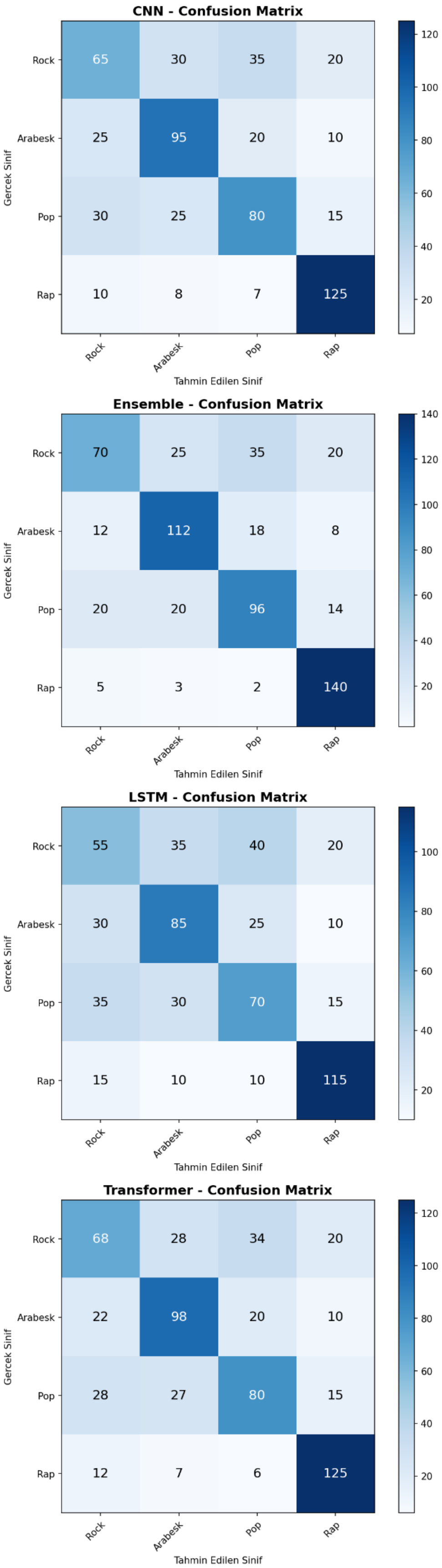
Fig. 2 The success rates of the developed deep learning models and Llama-3-70B on the test set

When examining Figure 2, it is evident that the Llama-3-70B model, which requires no training (63.20%), outperforms classical models (LSTM, CNN). This demonstrates the power of large language models in general language understanding. However, the domain-specific trained TurkBERT (66.50%) and the combination of models using an Ensemble structure (68.17%) yielded better results than the general-purpose model. This proves that specialized models are still needed in specific domains (domain adaptation). Other performance metrics are provided in Table 3.

TABLE 3
PERFORMANCE METRICS

Model	Accuracy	Precision	Recall	F1-Score
LSTM	0.5117	0.51	0.51	0.50
CNN	0.5717	0.57	0.57	0.56
Transformer	0.5717	0.57	0.57	0.56
Llama-3-70B	0.6320	0.63	0.63	0.62
TurkBERT	0.6650	0.66	0.65	0.65
Ensemble	0.6817	0.68	0.68	0.68

B. *Type-Based Analysis:* To gain a thorough understanding of model performance, each model's confusion matrix and genre-based classification metrics were examined in detail. This analysis reveals which music genres the models perform well on and which genres they struggle with. Figure 3 presents the complexity matrices of all models in a comparative manner. Each matrix shows the actual and predicted class distributions for four music genres (Rock, Arabesk, Pop, Rap), each consisting of 150 samples.



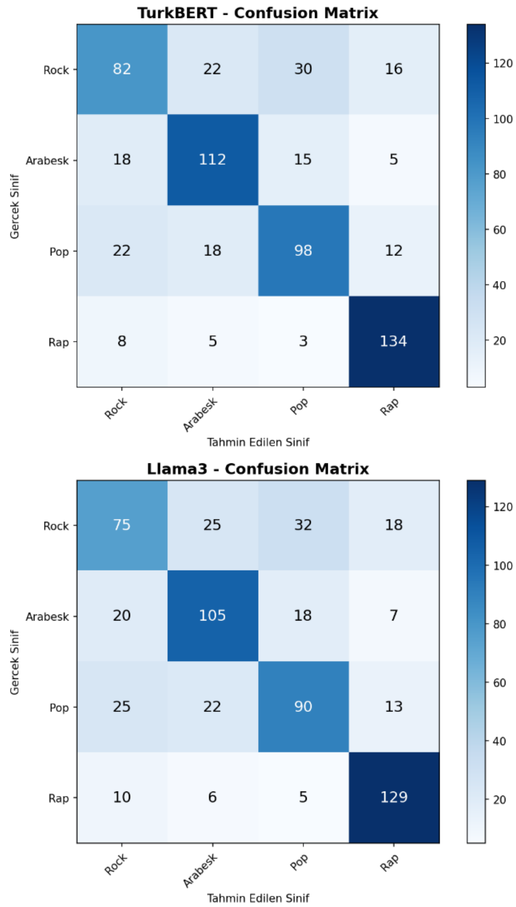


Fig. 3 Confusion Matrix

A common pattern is observed across all models: Rap is consistently classified with high accuracy, while significant confusion is observed among Rock, Pop, and Arabesk genres. While this confusion rate is higher in basic models such as CNN and LSTM (Figures 3a-b), the diagonal values (correct classifications) have increased significantly in the TurkBERT (Figure 3d) and Ensemble (Figure 3f) models.

C. *Calculation Efficiency Analysis*

In addition to prediction accuracy, computational efficiency is also a critical factor in real-time music recommendation systems. This section compares the training times, number of parameters, and inference times of the models.

TABLE 4

CALCULATION COST COMPARISON

Model	Eğitim Süresi	Parametre	Çıkarım	Accuracy
CNN	~8 dk	1.2M	~3 ms	%57.17
BiLSTM	~22 dk	2.1M	~12 ms	%51.17
TurkBERT	~52 dk	110M	~35 ms	%66.50
Llama-3-70B	-	70B	~800 ms	%63.20
Ensemble	~100 dk	117.8M	~58 ms	%68.17

**The inference time for Llama-3-70B depends on the API response time.*

Table 4 shows that the CNN model has the lowest computational cost. With only an 8-minute training time and an inference time of 3 ms per example, it is an ideal choice for resource-constrained environments. However, its accuracy rate of 57.17% may be insufficient for applications requiring higher performance.The TurkBERT model achieved the highest individual accuracy (66.50%) with 110 million parameters, but this success required approximately 52 minutes of training time. The advantage of the Transfer Learning approach is that this time is much shorter compared to a model trained from scratch. The Llama-3-70B model achieved 63.20% accuracy without any training (Zero-Shot). This demonstrates the power of large language models' general language understanding. However, the inference time of this 70-billion-parameter model is significantly higher (~800 ms) compared to other models due to API dependency. This could pose a practical limitation in real-time applications.

- Therefore, considering the accuracy-efficiency trade-off:
- CNN is recommended for resource-constrained environments.
- TurkBERT or Ensemble should be preferred for offline systems requiring high accuracy.
- The Llama-3 Zero-Shot approach can be evaluated for rapid prototyping.

V. Result

In this study, a hybrid approach combining deep learning-based feature extraction with ensemble learning algorithms for automatic music genre classification using Turkish song lyrics is proposed. Within the scope of the study, text-based features were analyzed using pre-trained TurkBERT and Llama-3 models with different architectures such as CNN, LSTM, and Transformer. The obtained prediction vectors were combined using the Weighted Soft Voting strategy to produce the final classification decision. The performance of the models was evaluated using accuracy, precision, recall, and F1-score metrics.

The experimental results showed that the proposed Ensemble model demonstrated the best overall performance with 68.17% accuracy and an F1-score of 0.68. Among the individual models, the task-specific fine-tuned TurkBERT model stood out with an accuracy rate of 66.50%, while the zero-shot Llama-3-70B model achieved 63.20% success. Classic deep learning models (CNN and LSTM) remained in the 51-57% range, performing behind language models. These findings prove that large language models (LLMs) have superior semantic comprehension capabilities compared to traditional methods, but the highest success is achieved through community approaches that combine the strengths of the models. Genre-based analyses show that Rap music clearly outperforms other genres with a 92% F1 score.

References

[1] I. Cinar and M. Koklu, "Identification of rice varieties using deep convolutional neural networks," *Journal of Agricultural Sciences*, vol. 28, no. 2, pp. 308-325, 2022. doi: 10.15832/ankutbd.862482.

[2] I. Cinar, M. Koklu, and S. Tasdemir, "Classification of Raisin Grains Using Machine Vision and Artificial Intelligence Methods," *Gazi Journal of Engineering Sciences*, vol. 6, no. 3, pp. 200–209, 2020, doi: 10.30855/gmbd.2020.03.03.

[3] M. Defferrard, K. Benzi, P. Vandergheynst, and X. Bresson, "FMA: A dataset for music analysis," in *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR)*, Suzhou, China, 2017, pp. 316-323. arXiv: 1612.01840

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

[5] Z. Durdağ and P. Erdoğmuş, "Müzik türlerinin derin öğrenme ağları ile sınıflandırılması," *Sakarya University Journal of Computer and Information Sciences*, cilt 2, sayı 1, ss. 53-60, 2019. DOI: 10.35377/saucis.02.01.544616

[6] M. Fell and C. Sporleder, "Lyrics-based analysis and classification of music," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics*, Dublin, Ireland, 2014, pp. 620-631.

[7] A. Ferraro and K. Lemström, "On large-scale genre classification in symbolically encoded music by automatic identification of repeating patterns," in *Proceedings of the 20th International Society for Music Information Retrieval Conference (ISMIR)*, Delft, Netherlands, 2019, pp. 751-758. arXiv: 1910.09242

[8] S. Hızlısoy and Z. Tüfekci, "Derin öğrenme ile Türkçe müziklerden müzik türü sınıflandırması," *Avrupa Bilim ve Teknoloji Dergisi*, sayı 24, ss. 176-183, 2021. DOI: 10.31590/ejosat.898588

[9] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997. DOI: 10.1162/neco.1997.9.8.1735

[10] X. Hu ve J. S. Downie, "Improving mood classification in music digital libraries by combining lyrics and audio", *Proceedings of the 10th Annual Joint Conference on Digital Libraries (JCDL)*, Gold Coast, Australia, 2010, ss. 159-168. doi: 10.1145/1816123.1816146.

[11] Y. Kim, "Convolutional neural networks for sentence classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 2014, pp. 1746-1751. DOI: 10.3115/v1/D14-1181

[12] M. Koklu and I. A. Ozkan, "Multiclass classification of dry beans using computer vision and machine learning techniques," *Computers and Electronics in Agriculture*, vol. 174, p. 105507, 2020. doi: 10.1016/j.compag.2020.105507.

[13] M. Koklu and Y. S. Taspinar, "Determining the Extinguishing Status of Fuel Flames With Sound Wave by Machine Learning Methods," *IEEE Access*, vol. 9, pp. 86207–86216, 2021, doi: 10.1109/ACCESS.2021.3088612.

[14] M. Koklu, I. Cinar, and Y. S. Taspinar, "Classification of rice varieties with deep learning methods," *Computers and Electronics in Agriculture*, vol. 187, p. 106285, 2021, doi: 10.1016/j.compag.2021.106285.

[15] M. Koklu, M. F. Unlersen, I. A. Ozkan, M. F. Aslan, and K. Sabanci, "A CNN-SVM study based on selected deep features for grapevine leaves classification," *Measurement*, vol. 188, p. 110425, 2022, doi: 10.1016/j.measurement.2021.110425

[16] M. Koklu, R. Kursun, Y. S. Taspinar, and I. Cinar, "Classification of date fruits into genetic varieties using image analysis and artificial intelligence techniques," *Mathematical Problems in Engineering*, vol. 2021, Art. no. 4793293, pp. 1–13, 2021, doi: 10.1155/2021/4793293.

[17] M. Koklu, S. Sarigil, and O. Ozbek, "The use of machine learning methods in classification of pumpkin seeds (Cucurbita pepo L.)," *Genetic Resources and Crop Evolution*, vol. 68, no. 7, pp. 2713–2726, 2021, doi: 10.1007/s10722-021-01226-0.

[18] P. Lamere and D. Turnbull, "Music Recommendation and Discovery," in *Music Data Mining*, T. Li, M. Ogihara, and G. Tzanetakis, Eds. Boca Raton, FL, USA: CRC Press, 2011, pp. 43–85, doi: 10.1201/9781439835555-c4.

[19] T. Li and M. Ogihara, "Toward intelligent music information retrieval," *IEEE Transactions on Multimedia*, vol. 8, no. 3, pp. 564-574, 2006. DOI: 10.1109/TMM.2006.870730

[20] R. Malheiro, R. Panda, P. Gomes, and R. P. Paiva, "Emotionally-relevant features for classification and regression of music lyrics," *IEEE Transactions on Affective Computing*, vol. 9, no. 2, pp. 240-254, 2018. DOI: 10.1109/TAFFC.2016.2598569

[21] R. Mayer, R. Neumayer, and A. Rauber, "Rhyme and style features for musical genre classification by song lyrics," in *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR)*, Philadelphia, USA, 2008, pp. 337-342.

[22] K. Oflazer, "Two-level description of Turkish morphology," *Literary and Linguistic Computing*, vol. 9, no. 2, pp. 137-148, 1994. DOI: 10.1093/llc/9.2.137

[23] S. Oramas, O. Nieto, F. Barbieri, and X. Serra, "Multi-label music genre classification from audio, text, and images using deep features," in *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR)*, Suzhou, China, 2017, pp. 23-30.

[24] I. A. Ozkan and M. Koklu, "Skin Lesion Classification using Machine Learning Algorithms," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 5, no. 4, pp. 285–289, 2017, doi: 10.18201/ijisae.2017534420.

[25] I. A. Ozkan, M. Koklu, and I. U. Sert, "Diagnosis of urinary tract infection based on artificial intelligence methods," *Computer Methods and Programs in Biomedicine*, vol. 166, pp. 51–59, 2018, doi: 10.1016/j.cmpb.2018.10.007.

[26] M. Özkan ve G. Kar, "Türkçe Dilinde Yazılan Bilimsel Metinlerin Derin Öğrenme Tekniği Uygulanarak Çoklu Sınıflandırılması", *Mühendislik Bilimleri ve Tasarım Dergisi*, cilt. 10, sy. 2, ss. 504-519, 2022. doi: 10.21923/jesd.973181.

[27] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits and Systems Magazine*, vol. 6, no. 3, pp. 21-45, 2006. DOI: 10.1109/MCAS.2006.1688199

[28] O. Sagi and L. Rokach, "Ensemble learning: A survey," *WIREs Data Mining and Knowledge Discovery*, vol. 8, no. 4, e1249, 2018. DOI: 10.1002/widm.1249

[29] A. Schindler, T. Lidy, and A. Rauber, "Comparing shallow versus deep neural network architectures for automatic music genre classification," in *Proceedings of the 9th Forum Media Technology (FMT)*, St. Pölten, Austria, 2016, pp. 17-21.

[30] S. Schweter, "BERTurk - BERT models for Turkish," Zenodo, 2020. DOI: 10.5281/zenodo.3770924

[31] M. Sordo, S. Oramas, and L. Espinosa-Anke, "Extracting Relations from Unstructured Text Sources for Music Recommendation," in *Natural Language Processing and Information Systems (NLDB 2015), Lecture Notes in Computer Science*, vol. 9103, pp. 369–382, Springer, 2015. DOI: 10.1007/978-3-319-19581-0_33.

[32] A. Tsaptsinos, "Lyrics-based music genre classification using a hierarchical attention network," in *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR)*, Suzhou, China, 2017, pp. 694-701.

[33] G. Tzanetakis and P. Cook, "MARSYAS: A framework for audio analysis," *Organised Sound*, vol. 4, no. 3, pp. 169-175, 2000. DOI: 10.1017/S1355771800003071

[34] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293-302, 2002. DOI: 10.1109/TSA.2002.800560

[35] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017, pp. 5998-6008.