T.C.
İSTANBUL ÜNİVERSİTESİ-CERRAHPAŞA
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

DOKTORA TEZİ

DERİN ÜRETİCİ MODELLER ile POLİFONİK MÜZİKAL İÇERİK ÜRETİMİ

Ahmet KAŞİF

DANIŞMAN
Doç. Dr. Selçuk SEVGEN

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği, Doktora Programı

Temmuz, 2024

TEZ KABUL VE ONAYI

Ahmet KAŞİF tarafından, **Doç. Dr. Selçuk SEVGEN** danışmanlığında hazırlanan **"DERİN ÜRETİCİ MODELLER ile POLİFONİK MÜZİKAL İÇERİK ÜRETİMİ"** başlıklı bu çalışma, jürimiz tarafından **24/07/2024** tarihinde yapılan sınav sonucunda **oy birliği** ile başarılı bulunarak **Doktora Tezi** olarak kabul edilmiştir.

Tez Jürisi

		İmza	Sonuç
DANIŞMAN	Doç. Dr. Selçuk SEVGEN İstanbul Üniversitesi-Cerrahpaşa Bilgisayar Mühendisliği Anabilim Dalı		☒ Kabul ☐ Ret
ÜYE	Doç. Dr. Sibel SENAN İstanbul Üniversitesi-Cerrahpaşa Bilgisayar Mühendisliği Anabilim Dalı		☒ Kabul ☐ Ret
ÜYE	Doç. Dr. Neyir ÖZCAN SEMERCİ Bursa Uludağ Üniversitesi Elektrik-Elektronik Mühendisliği Anabilim Dalı		☒ Kabul ☐ Ret
ÜYE	Prof. Dr. Ahmet SERTBAŞ İstanbul Üniversitesi-Cerrahpaşa Bilgisayar Mühendisliği Anabilim Dalı		☒ Kabul ☐ Ret
ÜYE	Dr. Öğr. Üyesi Oğuz ALTUN Yıldız Teknik Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı		☒ Kabul ☐ Ret

BEYAN

Bu tez çalışmasının kendi çalışmam olduğunu, tezin planlanmasından yazımına kadar bütün aşamalarda etik dışı davranışımın olmadığını, bu tezdeki bütün bilgileri akademik ve bilimsel etik kuralları içinde elde ettiğimi, bu tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara kaynak gösterdiğimi ve bu kaynakları da kaynaklar listesine aldığımı, yine bu tezin çalışılması ve yazımı sırasında patent ve telif haklarını ihlal edici bir davranışımın olmadığını ve her türlü hukuki sorumluluğu aldığımı kabul ederim.

Ahmet KAŞİF

(İmza)

Aileme ithaf ediyorum…

BÜTÇE DESTEKLERİ

DERİN ÜRETİCİ MODELLER ile POLİFONİK MÜZİKAL İÇERİK ÜRETİMİ

Bu tez çalışması için herhangi bir kurumdan bütçe desteği alınmamıştır.

TEŞEKKÜR

Doktora araştırmalarım süresince rehberlik ve destekleri için danışmanım Doç. Dr. Selçuk SEVGEN'e içten şükranlarımı sunarım. Cesaretimi artıran, sabrı ve çalışkanlığı tavsiye ederek zorlu araştırma sürecinde her zaman yanımda olan Doç. Dr. Selçuk SEVGEN hocama yol göstericiliği için minnettarım.

Aileme ve arkadaşlarıma sevgi ve destekleri için teşekkür ederim. Her koşulda yanımda olan annem, babam ve kardeşim, bu tez çalışmasının ortaya çıkmasında en büyük paylardan birine sahip oldular. Tez çalışmamı kendilerine ithaf ediyorum.

Doktora sürecimde birlikte çalışma ve deneyimlerinden yararlanma fırsatı sundukları için Prof. Dr. Çağatay ÇATAL'a ve Doç. Dr. Alper ÖZCAN'a teşekkür ederim. Doktora tezimin gelişim süreci boyunca değerli katkıları ve geri dönüşleri ile beni destekleyen Doç. Dr. Sibel SENAN'a ve Doç. Dr. Neyir ÖZCAN SEMERCİ'ye teşekkür ederim. Bu doktora çalışmasını gerçekleştirirken Bursa Teknik Üniversitesi'nde birlikte mesai yaptığım çalışma arkadaşlarıma teşekkür ederim. Sağladıkları araştırma ortamı ile akademik ve kişisel gelişimime katkı sağlayan tüm hocalarıma minnettarım.

Temmuz 2024 Ahmet KAŞİF

İÇİNDEKİLER

Sayfa No

TEZ KABUL VE ONAYI ..... **ii**
BEYAN ..... **iii**
BÜTÇE DESTEKLERİ ..... **v**
TEŞEKKÜR ..... **vi**
İÇİNDEKİLER ..... **vii**
ŞEKİL LİSTESİ ..... **ix**
TABLO LİSTESİ ..... **xi**
SİMGE VE KISALTMA LİSTESİ ..... **xii**
ÖZET ..... **xiii**
ABSTRACT ..... **xv**
1. GİRİŞ **1**
2. KAVRAMSAL ÇERÇEVE **4**
2.1. Müzikal İçeriklerde Gösterim 5
2.1.1. Sinyal Gösterim 5
2.1.2. Sembolik Gösterim 8
2.1.3. Çizge Gösterimi 10
2.2. Uzun-Vadeli Bağımlılıklar Problemi 12
2.3. Yinelemeli Sinir Ağları 14
2.4. Uzun-Kısa Dönem Hafıza Sinir Ağları 15
2.5. Öz-dikkat Ağları (Self-Attention) 17
2.6. Çok-başlıklı Dikkat Ağları (Multi-Head Attention) 18
3. YÖNTEM **20**
3.1. Veri Setleri ve Ön İşleme 20
3.2. MTHA-LSTM 22
3.2.1. Veri Gösterimi 22
3.2.2. Mimari Tasarım 23
3.3. GAMA-DL 27
3.3.1. Veri Gösterimi 27

3.3.2. Mimari Tasarım 31
3.4. Eşlik Üretimi 37
4. BULGULAR 40
4.1. Nicel Ölçüm Yöntemleri 40
4.2. Nitel Ölçüm Yöntemleri 42
4.3. MTHA-LSTM Sonuçları 45
4.3.1. Nicel Ölçümler 46
4.3.2. Nitel Ölçümler 48
4.4. GAMA-DL Sonuçları 49
4.4.1. Nicel Ölçümler 49
4.4.2. Nitel Ölçümler 51
4.5. Hafıza Karmaşıklığı Analizi 53
5. TARTIŞMA 55
6. SONUÇ ve ÖNERİLER 57
KAYNAKLAR 59
EKLER 65
İNTİHAL RAPORU İLK SAYFASI 66
ETİK KURUL İZİN YAZISI 67
KURUM İZNİ YAZILARI 68
ÖZGEÇMİŞ 69

ŞEKİL LİSTESİ

Sayfa No

Şekil 2.1: Sembolik gösterim ile tek-sesli, çok-sesli ve çok-enstrümanlı çok-sesli müzik 4

Şekil 2.2: Sinyal gösterimde (a) Genlik/Zaman, (b) Genlik/Frekans, (c) Spektrogram temsilleri 8

Şekil 2.3: Espana Opus 165 Prelude (Albeniz) parçası (a), piano-roll gösterimi (b), MIDI-like gösterimi (c) 10

Şekil 2.4: Hiyerarşik modelleme ile müzikal içeriğin gösterimi 11

Şekil 2.5: Uzun-kısa dönem hafıza (LSTM) hücresi 16

Şekil 3.1: Kayar pencere yöntemi ile girdi vektörü - hedef değişken eldesi 22

Şekil 3.2: MTHA-LSTM Mimari Tasarım 25

Şekil 3.3: Çizgelerde komşulukların gösterimi, (a) Çizge örneği, (b) Komşuluk matrisi gösterimi, (c) Komşuluk listesi gösterimi 28

Şekil 3.4: Sembolik Gösterim ile Müzikal Ölçü İlişkileri [51] 29

Şekil 3.5: Müzikal ölçü ilişkilerinin çizge veri yapısı ile gösterimi 30

Şekil 3.6: Müzikal ölçü koşullu nota seviye gösterim 31

Şekil 3.7: Melodi üretici modül mimari yapısı 32

Şekil 3.8: Müzikal ölçü üretici modül mimari tasarımı 34

Şekil 3.9: Müzikal ölçü koşullu nota üretici modül mimari tasarımı 36

Şekil 3.10: Müzikal ölçü koşullu nota üretici modül mimari tasarımı (Devamı) 37

Şekil 3.11: Melodi ve eşlik nota dizileri ile sembolik gösterimde temsili müzikal içerik 39

Şekil 4.1: MTHA-LSTM kullanıcı araştırması soru örneği 43

Şekil 4.2: GAMA-DL kullanıcı araştırması soru örneği 44

Şekil 4.3: Model Eğitim Aşaması Kayıp Eğrileri 47

Şekil 4.4: Kosinüs benzerlik matrisinde ölçüler arası benzerlik ortalamaları 50

Şekil 4.5: GAMA-DL Kullanıcı Çalışması Sonuçları (Tümü) 52

Şekil 4.6: GAMA-DL Kullanıcı Çalışması Sonuçları (Tecrübeli)..... 53

Şekil 4.7: GAMA-DL Kullanıcı Çalışması Sonuçları (Sıradan Dinleyici) 53

TABLO LİSTESİ

Sayfa No

Tablo 3.1: KPM ve MAESTRO veri setlerinin istatistiki özellikleri 21

Tablo 4.1: YZ model öğrenme başarımları 46

Tablo 4.2: MTHA-LSTM nicel ölçüm sonuçları..... 48

Tablo 4.3: MTHA-LSTM Kullanıcı Araştırması Sonuçları..... 49

Tablo 4.4: Varyans ve ÖDS metrikleri ile nicel sonuçlar 50

Tablo 4.5: Yöntemlerin parametre sayıları..... 54

Tablo 5.1: GAMA-DL ve MTHA-LSTM yöntemlerinin literatürdeki çalışmalar ile karşılaştırması..... 56

SİMGE VE KISALTMA LİSTESİ

Simgeler **Açıklama**

M : Müzik parçasındaki motiflerin kümesi
L : Pencere boyutu
|m| : Bir motifin uzunluğu
M : Eşlik üretimi modülünde melodik nota dizisi
n : Eşlik üretimi modülünde nota-akor sözlüğü
Z : Eşlik üretimi modülünde eşlik akor dizisi
b : Eşlik üretimi modülünde baz melodik nota
c : Eşlik üretimi modülünde seçilen eşlik akoru
p : Kayar pencere yönteminde pencere uzunluğu

Kısaltmalar **Açıklama**

DÖ : Derin Öğrenme
GAMA-DL : Çizge Destekli Müzikal Dikkat Ağı (Graph Reinforced Music Attention)
KB : Kosinüs Benzerliği
LSTM : Uzun-Kısa Vade Hafıza
MIR : Müzikal Bilgi Erişim (Musical Information Retrieval)
MTHA : Çok-Başlıklı Dikkat Ağı (Multi-Head Attention)
ÖDS : Özgün Değer Sayısı
RNN : Yinelemeli Sinir Ağı
TİS : Tekrarlı İfade Sayısı
ÜYZ : Üretici Yapay Zeka
YZ : Yapay Zeka

ÖZET

DOKTORA TEZİ

DERİN ÜRETİCİ MODELLER İLE POLİFONİK MÜZİKAL İÇERİK ÜRETİMİ

Ahmet KAŞİF

İstanbul Üniversitesi-Cerrahpaşa

Lisansüstü Eğitim Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği, Doktora Programı

Danışman: Doç. Dr. Selçuk SEVGEN

Bu tez çalışmasında derin üretici ağlar ile sembolik polifonik (çok-sesli) müzikal içerik üretimi, Batı Klasik Tonal Müziği bilgi uzayında gerçeklenmiştir. Çok-sesli müzikal içerik üretimi, tek-sesli melodi üretimi ve tek-sesli eşlik üretimi olarak iki alt-problem temelinde incelenmiştir. Sembolik sıralı veri gösteriminin kullanıldığı çalışmada, Yapay zekâ ile melodi üretimi alandaki en önemli teknik problemlerden uzun-dönem ilişkisel yapının korunumu probleminin aşılmasına çalışılmış ve (MTHA-LSTM) ile (GAMA-DL) yöntemleri geliştirilmiştir. MTHA-LSTM, çok-başlıklı dikkat ağları (MTHA) ile uzun-kısa vade hafıza (LSTM) ağlarını birlikte kullanan iki katmanlı ve hiyerarşik bir derin öğrenme yöntemidir. GAMA-DL, yüksek seviyeli müzikal ölçü üretimi ve düşük seviyeli nota üretimi olmak üzere iki aşamaya sahip derin öğrenme yönteminin genel adıdır. Geliştirilen yöntemler nicel ölçüm yöntemleri ve kullanıcı araştırmaları ile sınanmıştır. Elde edilen sonuçlar, uzun-vadeli bağımlılıklar problemini aşarak daha uzun ve yüksek kaliteli müzikal içerik üretmede önerilen yöntemlerin etkinliğini göstermektedir.

Temmuz 2024 , 83 sayfa.

Anahtar kelimeler: sembolik müzik, müzik üretimi, derin öğrenme, üretici yapay zekâ, hiyerarşik müzik üretimi, melodi üretimi,

ABSTRACT

Ph.D. THESIS

GENERATING POLYPHONIC MUSICAL CONTENT WITH DEEP GENERATIVE MODELS

Ahmet KAŞİF

İstanbul University-Cerrahpaşa

Institute of Graduate Studies

Department of Computer Engineering

Computer Engineering Ph.D Programme

Supervisor : Assoc. Prof. Dr. Selçuk SEVGEN

In this thesis, the generation of symbolic polyphonic musical content using deep generative networks has been implemented within the knowledge domain of Western Classical Tonal Music. The production of polyphonic musical content has been examined based on two sub-problems: the generation of monophonic melodies and the generation of monophonic accompaniment. In the study, which uses symbolic sequential data representation, efforts have been made to overcome the problem of preserving long-term relational structure, one of the most significant technical challenges in the field of artificial intelligence-driven melody generation. Consequently, the methods (MTHA-LSTM) and (GAMA-DL) have been developed. MTHA-LSTM is a two-layered hierarchical deep learning method that employs multi-headed attention networks (MTHA) in conjunction with long short-term memory (LSTM) networks. GAMA-DL refers to the general framework of a deep learning method that operates in two stages: high-level musical measure generation and low-level note generation. The developed methods have been tested through quantitative measurement methods and user studies. The obtained results demonstrate the effectiveness of the proposed methods in

producing longer and higher-quality musical content by overcoming the long-term dependency problem.

July 2024, 83 pages.

Keywords: symbolic music, music production, deep learning, generative artificial intelligence, hierarchical music production, melody generation

1. GİRİŞ

Müzikal içerik üretimi, sesleri bir araya getirerek müzikal bir eser oluşturma sanatıdır. Bilgisayar destekli müzikal içerik üretiminin tarihi ise ilk ses kayıt cihazlarının icadına kadar uzanan uzun ve karmaşık bir geçmişe sahiptir. Bilgisayar destekli müzikal içerik üretiminin ilk örnekleri, 19. yüzyılın sonlarında Thomas Edison'un fonografı gibi cihazlarla kaydedilen basit sesler olarak ortaya çıkmıştır [1]. 20. yüzyılda, manyetik bant ve ses mikseri gibi yeni teknolojilerin geliştirilmesi, müzik prodüksiyonunun daha karmaşık ve sofistike hale gelmesini sağlamıştır. Günümüzde, dijital teknolojinin gelişmesi, her zamankinden daha fazla insanın müzik prodüksiyonu yapabilmesine ve kendi müziklerini yaratabilmesine olanak sağlamaktadır ve müzik popüler kültürün önemli parçalarından biri olma özelliğini korumaktadır.

Müzikal Bilgi Erişim (MIR) veya Hesaplamalı Müzik bilim alanı, müzik, psikoloji, sinyal işleme ve bilişim gibi farklı alanları bir arada barındırmakta ve gelişen teknolojinin müzik dünyasındaki yansımasını disiplinler arası bir bakış açısıyla temsil etmektedir [2]. MIR kapsamında sınıflandırma, karar destek sistem tasarımı, otonom yapay müzik transkripsiyonu ve müzikal içerik üretimi gibi alt problemler günümüzde oldukça sık çalışılan araştırma problemleri olarak karşımıza çıkmaktadır. Müzik dünyasında tam-otomatik veya insan-destekli yapay zekâ çözümlerinin kullanımının, müzikal kabiliyetinin geliştirilmesi ve çeşitlendirilmesinde önemli katkılar sağladığı ve bu katkıların geliştirilmesine ihtiyaç bulunduğu görülmektedir [3].

Müzik üretimi, içeriğin sanatsal yapısı sebebiyle günümüzde halen çoğunlukla insan kaynaklı bir süreç olarak gelişimini sürdürmektedir. Profesyonel bir besteci olmanın geleneksel yolu uzun ve zorlu şartlar sunmakta, aday bestecinin uzun yıllar eğitim ve mentorluk almasını gerektirmektedir. Bununla birlikte gelişen teknoloji ve değişen kullanıcı talepleri ile elektronik yayınların, video oyunlarının ve müziğin kullanımını gerektiren diğer medya araçlarının çoğalması, müziğe olan talebi sürekli artırmakta ve çeşitlendirmektedir [4]. Modern hesaplama araçlarının sunduğu kabiliyetler, yapay zekânın son yirmi yılda önemli bir gelişim göstermesinin önünü açmıştır. Özellikle son on yılda ise yapay zekâ ile görsel, metinsel veya sinyal tabanlı yapay içerik üretimi, popüler araştırma konuları olarak öne çıkmaktadır. Yapay müzik üretimi de bu alanlar arasında teknik açıdan doğru ve müzikal açıdan zengin ve etkileyici

müzikal içerikleri yapay olarak üretme görevinin karşılığı olarak ifade edilmekte ve özgün bir zorluk ortaya koymaktadır [5][6].

Müzikal içerikler ile gerçekleştirilen çalışmalarda kullanılan en popüler veri gösterimleri sembolik ve sinyal gösterimlerdir [6][7]. Her iki temsil de verileri sıralı yapıda incelemektedir. Sıralı verilerin temsilindeki en temel teknik problemlerden biri uzun-vadeli bağımlılıklar problemidir [8]. Uzun vadeli bağımlılıklar problemi, sıralı verilerde geçmişte görülen bir bilginin daha sonraki değerlere etkisinin temsilinde yaşanan güçlüktür. Bu güçlük, Yapay Zekâ (YZ) modellerinin sıralı verilerde değişkenler arasındaki ilişkileri unutması ve uzun dizilerdeki tahminlerin olumsuz etkilenmesi ile gözlemlenmektedir. Müzikal içerik üretimi bağlamında bu problem, YZ modelleri ile elde edilen müzikal içerik uzunluğunu sınırlamakta ve müzikal kaliteyi düşürmektedir [9].

Bu tez çalışmasında, çok-sesli (polifonik) müzikal içerik üretimi tek-sesli (monofonik) melodi üretimi ve tek-sesli eşlik üretimi (akompani) olmak üzere iki farklı alt problem nezdinde incelenmiştir. Tek-sesli melodi üretimi probleminin çözümü için iki yeni yöntem geliştirilmiştir. Bu yöntemlerin ilki, çok-başlıklı dikkat ağları (MTHA) ile uzun-kısa vade hafıza (LSTM) ağlarını birlikte kullanan iki katmanlı ve hiyerarşik bir derin öğrenme yöntemidir (MTHA-LSTM). Diğer yöntem ise çizge koşullu müzik dikkat (Graph conditional music attention - GAMA-DL) yöntemidir. GAMA-DL, nota üretimini müzikal ölçü (measure), seviyesindeki ilişkilerle destekleyerek müzikal ölçü koşullu nota üretimi gerçekleştirmektedir. Tek-sesli müzikal melodilere eşlik eklenerek çok-sesli müzikal içerik elde edilmesi amacıyla da tek-sesli eşlik üreteci geliştirilmiştir. Tez araştırması kapsamında elde edilen katkılar şunlardır:

- Çok-sesli müzikal içerik üretimi problemi, tek-sesli melodik içerik üretimi ve tek-sesli eşlik üretimi olmak üzere iki alt-problem nezdinde incelenmiştir. Elde edilen nihai eser, eşlik ile zenginleştirilmiş melodik bir ezgiyi ifade eden çok-sesli müzikal bir içeriktir.

- Tek-sesli melodik içerik üretimi probleminde çok-başlıklı dikkat ağları ile uzun-kısa vadeli hafıza (MTHA-LSTM) ağlarının birlikte kullanımı ile uzun-vadeli ilişkileri güçlü bir şekilde temsil edebilen ÜYZ yöntemi geliştirilmiştir.

- Müzikal ölçü ve nota seviyesinde birbirleri ile entegre çalışan üretici ağların melodi üretimi probleminde kullanıldığı, ölçü-koşullu müzik-dikkat ÜYZ yöntemi (GAMA-DL)

geliştirilmiştir. Yöntem, yüksek uzun-vadeli ilişki temsil kabiliyeti, düşük hafıza karmaşıklığı ve yüksek müzikal kaliteyi göstermektedir.

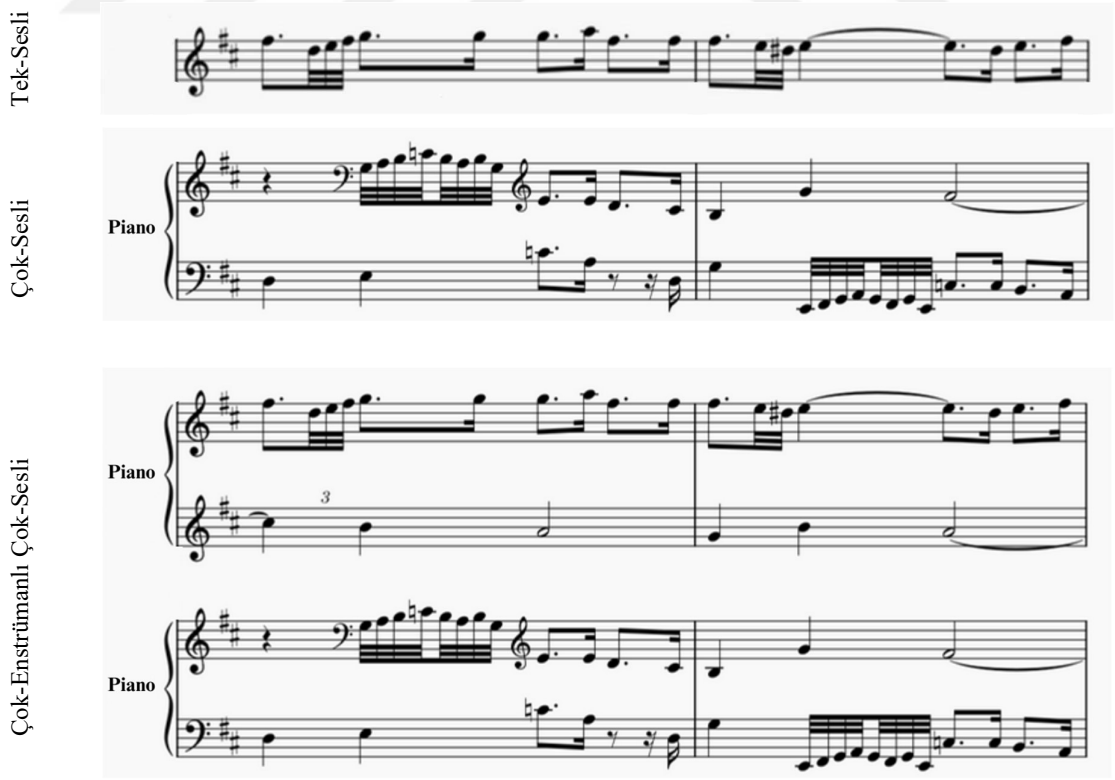
- MTHA-LSTM yönteminde kullanılan dikkat ağlarının hafıza karmaşıklığını iyileştirmek ve melodi üretiminde müziğin hiyerarşik özelliklerinin etkisini artırabilmek amacıyla müzikal ölçü seviyesinde ve nota seviyesinde iki farklı gösterim kullanılmıştır. Sembolik müzikal içerikler müzikal ölçü seviyesinde (yüksek-seviye) çizge gösterimi ile, nota seviyesinde (düşük-seviye) ise sıralı gösterim (sequential) ile temsil edilmiştir.

Tez, giriş, kavramsal çerçeve, yöntem, bulgular, tartışma ve öneriler olmak üzere beş bölümden oluşmaktadır. Giriş bölümünde, araştırmanın konusu, amacı, araştırma soruları ve hipotezler ile yöntemi hakkında bilgi verilmektedir. Kavramsal çerçeve bölümünde, müzikal bilginin gösterimi, uzun-vadeli bağımlılıklar problemi ve müzikal içerik üretiminde kullanılan derin öğrenme (DÖ) yöntemleri hakkında detaylı bilgi sunulmaktadır. Yöntem bölümünde, kullanılan veri setleri, önerilen MTHA-LSTM ve GAMA-DL yöntemleri için veri gösterimleri ve deney tasarımları açıklanmaktadır. Bulgular bölümünde, nicel ve nitel ölçümlerden elde edilen bulgular sunulmakta ve önerilen yöntemlerin etkinliği literatürdeki çalışmalar ile karşılaştırılmaktadır. Tartışma ve öneriler bölümünde, bulgular yorumlanmakta ve literatürle ilişkilendirilmektedir.

2. KAVRAMSAL ÇERÇEVE

Müziğin doğası içerdiği enstrüman sayısı ve belirli bir anda içerdiği nota sayısına göre tek sesli monofoni, tek sesli polifoni ve çok sesli polifoni olarak 3 alt sınıfta incelenmektedir [7][10]. Tek enstrümanlı monofoni, en temel doku olarak, tek bir enstrüman tarafından çalınan tek bir melodik çizgiyi sunmaktadır. Tek enstrümanlı polifoni ise tek bir enstrümanda birden fazla melodik çizgiyi aynı anda örerek bir karmaşıklık katmanı eklemektedir. Çok enstrümanlı polifoni de her biri farklı melodik çizgiler sunan birden fazla enstrümanın ses potansiyelini açığa çıkararak karmaşık müzikal dokuyu temsil etmektedir [11].

Müziğin doğasının farklı sınıflarda sembolik gösterim kullanılarak incelemesi, J. Sebastian Bach'ın bestelediği "Prelude and Fugue in C minor BWV 847" başlıklı parça özelinde Şekil 2.1'de sunulmaktadır.



Şekil 2.1: Sembolik gösterim ile tek-sesli, çok-sesli ve çok-enstrümanlı çok-sesli müzik

Aynı anda çalınan nota sayısı ile kullanılan enstrüman sayısındaki artış, müzikal karmaşıklığı artırdığı gibi, matematiksel temsili de zorlaştırmaktadır. Tek-sesli müzikal içerik bilgi uzayları, literatürde müzikte melodik ve ritmik yapının modellenmesi için kullanılmakta, çok-sesli müzikal içerik bilgi uzayları ise melodi ve ritmin yanı sıra eşlik gibi yapıların modellenmesinde kullanılmaktadır. Tez çalışması kapsamında tek-enstrümanlı çok-sesli müzikal içerik bilgi uzayının, üretici yapay zekâ yöntemleri ile analiz edilmesi hedeflenmiştir. Enstrümanlar arası ilişkilerin modellenmesi, problemin araştırma uzayını oldukça genişleten ve çözüm karmaşıklığını artıran bir zorluk oluşturması nedeniyle tercih edilmemiştir.

2.1. Müzikal İçeriklerde Gösterim

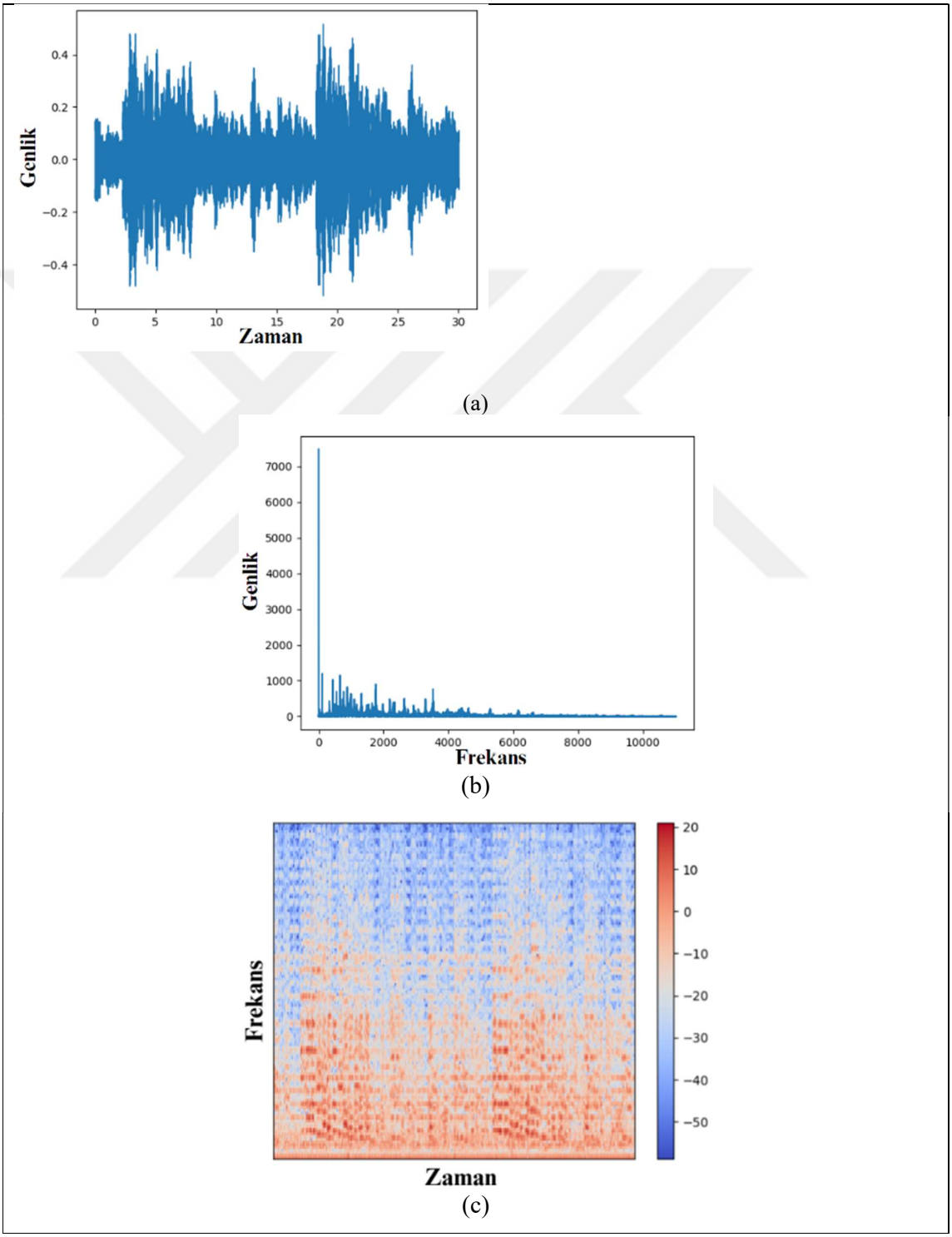
Gösterimler müzikal içeriğin vektör ve matris gibi temel matematiksel yöntemlerle modellenirken kullanılacak temsilleri ifade etmektedir. Literatürde problem odaklı özel çözümler dışında gösterim türleri ağırlıklı olarak sembolik ve sinyal gösterimler olarak iki temel sınıfta incelenmektedir. Sembolik gösterim ile gerçekleştirilen müzikal analizlerin genellikle ayrık değişkenler kullandığı, sinyal gösterimlerde ise sürekli değişkenlerin tercih edildiği görülmektedir.

Bununla birlikte, her iki gösterimde gerçekleştirilen çalışmaların, müziği sadece matematiksel uzaya aktararak modellemeye çalışan bir yaklaşım ortaya çıkardığı anlaşılmaktadır [12]. Sinyal gösterimlerin spektrogram veya kromagram dönüşümleri ile desteklenerek kullanılması, sembolik gösterimlerin çizge ve ağaç gibi gösterimlerle desteklendiği çalışmalar, alandaki bu alt-problemlerin yeni yaklaşımlarca gelişim gösterdiğini gözler önüne sermektedir [13]. Hiyerarşik ve çok boyutlu özellikleri bulunan müzikal bilgi uzayı, bu özellikleri öne çıkarılarak temsil edildiğinde daha gerçekçi ve kapsamlı bir araştırma ortaya konulabileceği anlaşılmaktadır. Bu bölümde, sinyal gösterim ve sembolik gösterim hakkında temel bilgiler sunulmakta, ardından çizge gösterimi, sembolik temsil tabanlı bir geliştirme olarak sunulmaktadır.

2.1.1. Sinyal Gösterim

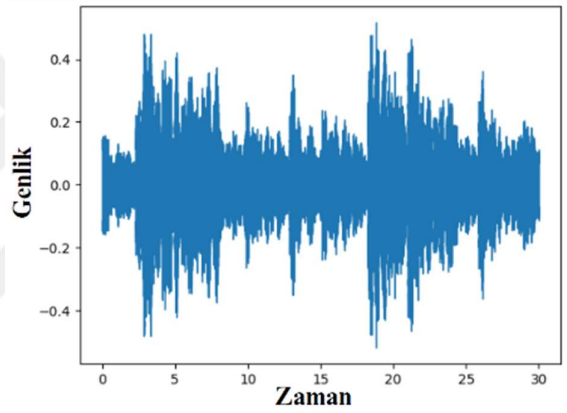
Sinyal gösterimi, müzikal sesleri ve olayları zamana bağlı bir şekilde temsil eden matematiksel bir temsildir. Bu gösterimde ses dalgalarının genlik ve frekans gibi fiziksel özellikleri sayısal değerlere dönüştürülerek müzikal bilgi kodlanmaktadır. Müzikal bilginin sinyal gösterimde

kullanıldığı yöntemler arasında ses dalga formu, spektrogramlar ve fourier dönüşümü yöntemleri öne çıkmaktadır (

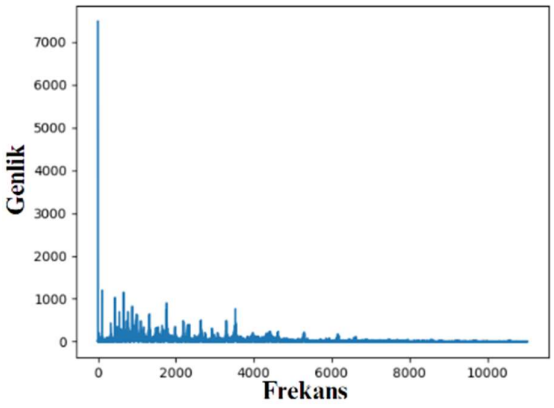


Şekil 2.2). Ses dalga formları, ses dalgalarının zamana bağlı genlik değişimlerini gösteren grafiksel bir temsil ortaya koymaktadır. Spektrogramlar, ses dalgalarının frekans ve zaman

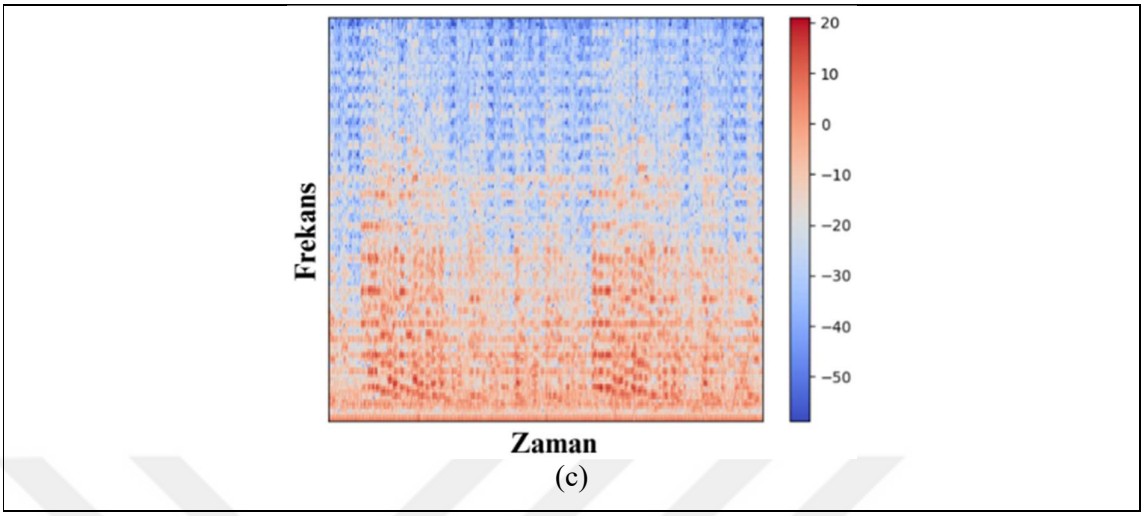
içerisindeki enerji dağılımını gösteren görsel bir temsil sunmakta, müzikal içerikleri görüntü verileri haline getirmektedir. Fourier dönüşümü ise ses dalgalarını sinüs ve kosinüs dalgalarının toplamı olarak ifade eden matematiksel bir işlem sunmaktadır. Sinyal gösterimlerin ortak özelliği, düşük seviye ses özniteliklerinin kullanılması vesilesi ile, bilgi çıkarımını yüksek karmaşıklık içeren bir araştırma problemi olarak incelemeleridir. Bu durumun en önemli çıktısı ileri derecede müzikal keskinliğe sahip içerikler elde edilebilmesidir. Bununla birlikte yüksek hesaplama karmaşıklığı gereksinimi bulunması ve içeriklerin düşük seviyeli özelliklerinin insanlar tarafından yorumlanmasında ortaya çıkan zorluklar, mevcut çalışmaların önündeki en önemli problemler olarak durmaktadır [4][7].



(a)



(b)

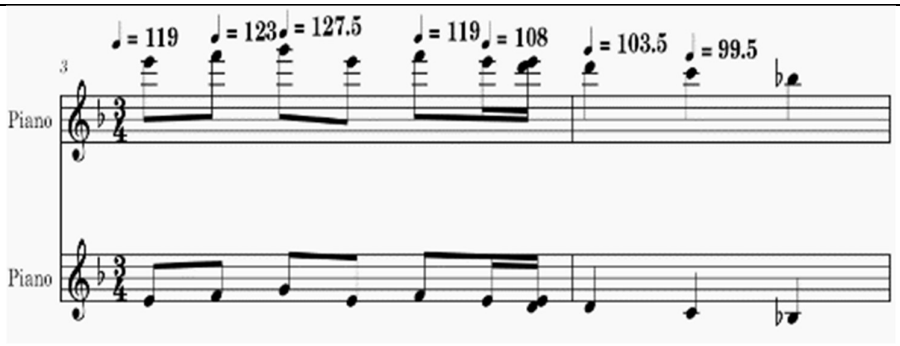


Şekil 2.2: Sinyal gösterimde (a) Genlik/Zaman, (b) Genlik/Frekans, (c) Spektrogram temsilleri

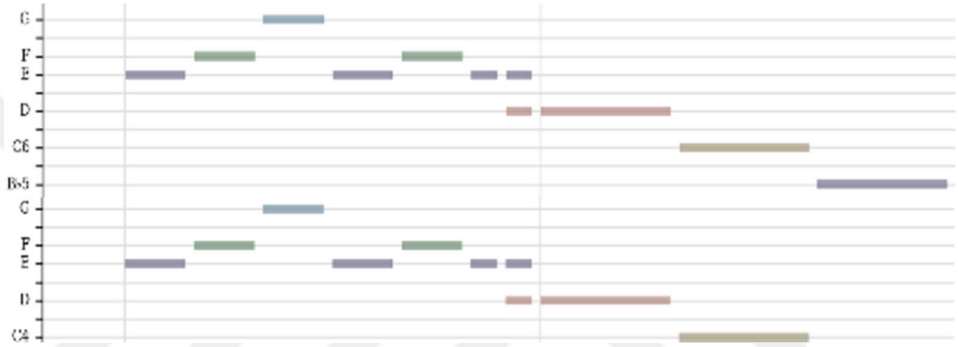
2.1.2. Sembolik Gösterim

Sembolik gösterim, müzikal notaları, ritimleri ve diğer müzikal unsurların, özel semboller ve kurallar kullanılarak temsil edildiği bir gösterim yöntemidir. Gösterim müzik teorisinde ve nota yazımında yaygın olarak kullanılmakta, sinyal gösterim ile karşılaştırıldığında nota, çalım süresi, ses seviyesi gibi daha yüksek seviyeli değişkenler ortaya çıkarmaktadır. Sembolik müzik içeriklerinin temel bileşenleri nota, sus ve akorlar gibi üst düzey özelliklerdir. Notalar ve akorlar belirli müzikal dizilerde bir araya gelerek müzik parçalarını oluşturmaktadır [14].

Dijital teknolojilerin gelişmesi ile birlikte, makine tarafından okunabilen çok sayıda müzik formatı geliştirilmiştir. Bunların en önde geleni ise MIDI (Müzik Enstrümanı Dijital Arayüzü) formatıdır [15]. MIDI formatı, sınırlı sayıda sembolü müzikal alfabe kabul ederek, müziği sembolik gösterimi aracılığıyla dijital yapıda işlemeyi mümkün kılmaktadır. Bir MIDI dosyası, içeriğindeki farklı notalar ve süreler için bilgiyi, önceden tanımlanmış bir zaman ölçüsü (bir çeyrek notanın alt bölümü) ile sayısal mesajlar halinde kodlamaktadır. Bu format, kompakt bir müzik temsili sağladığı için hesaplamalı müzik araştırmalarında yaygın olarak kullanılmaktadır [16], [17].



(a)

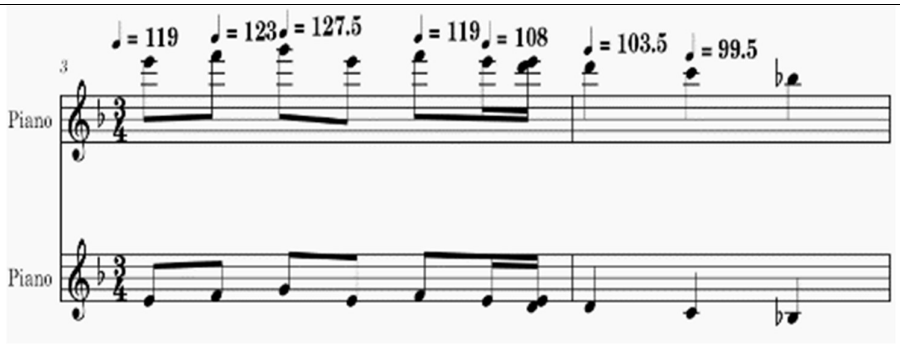


(b)

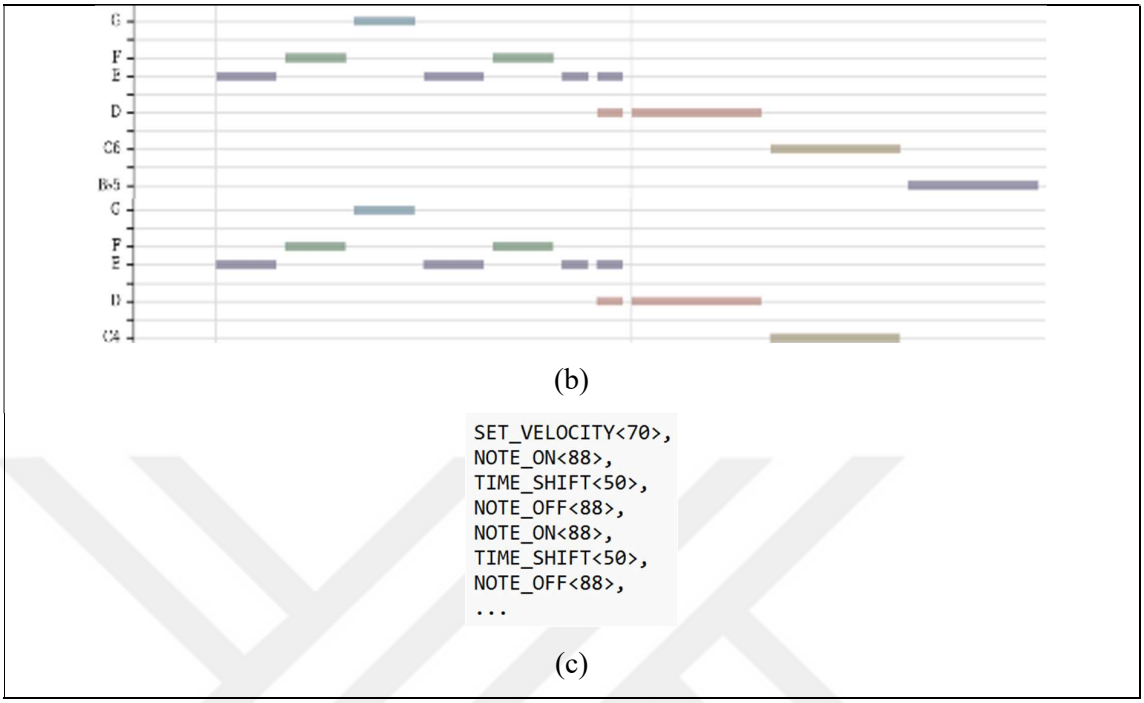
```
SET_VELOCITY<70>,
NOTE_ON<88>,
TIME_SHIFT<50>,
NOTE_OFF<88>,
NOTE_ON<88>,
TIME_SHIFT<50>,
NOTE_OFF<88>,
...
```

(c)

Şekil 2.3'te bir MIDI parçanın piano-roll ve MIDI-like sembolik gösterimleri birlikte sunulmaktadır.



(a)



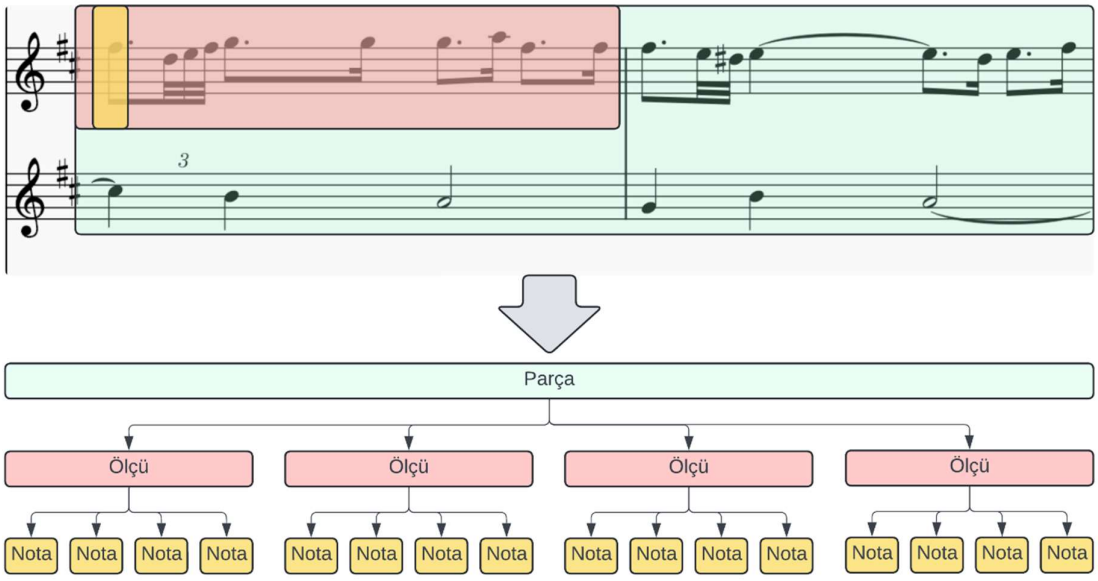
Şekil 2.3: Espana Opus 165 Prelude (Albeniz) parçası (a), piano-roll gösterimi (b), MIDI-like gösterimi (c)

2.1.3. Çizge Gösterimi

Çizgeler, karmaşık hiyerarşik ilişkileri modellemek için esnek bir gösterim sunmaktadır. Çizge temsilleri, zaman-serisi gösterimindeki katı doğrusal diziler yerine müzikal düğümler ve bu düğümlerin aralarındaki ilişkilerin daha net bir ilişki ağı ile temsilini mümkün kılarak son yıllarda popüler sembolik temsillerden biri haline gelmiştir. Çizgeler MIR alanında nota dizilimleri arasındaki ilişkileri belirlemek, nota-akor ilişkilerini incelemek, yapay içerik üretimi ve müzik sınıflandırma gibi birçok araştırma probleminde kullanılmaktadır.

Jeong ve ark. sembolik müzikal içerik kullanarak insan tarafından çalınan piyano performansına benzeyen bir ses üretme amacıyla çizge gösterimlerini kullanmıştır [18]. Gösterimde notalar çizgelerdeki düğümlere, kenarlar ise notalar arasındaki ilişkilere karşılık kullanmış, çok-sesli müzikal içeriklerin hem zamana-dayalı ilişkileri hem de akor ilişkileri tek bir temsilde ifade edilebilmiştir. Karystinaios ve Widmer de nota ve susları düğüm; çalım süresi ve çalım süresini ise kenar bilgisi olarak kullanarak çizge gösterimi elde etmiş, klasik müzikte kadans tespitini (tonal merkez değişikliklerini belirlemek) amaçlamıştır [19]. Guo ve Liu ise müzik öneri motoru hazırlama amacıyla heterojen müzik çizge gösterimini önermiştir [20].

Müzik, notalardan oluşan ve belirli bir düzen içinde icra edilen bir sanattır. Bu düzen, notalar, ölçüler ve parçalar gibi farklı seviyelerdeki gösterimler sağlanmakta ve bu gösterimler arasında hiyerarşik ilişkiler bulunmaktadır. Tez çalışmasında sembolik müzikal içerik uzayı nota seviyesinde nota, çalım süresi ve ses seviyesi olmak üzere üç değişkenden oluşan sembolik veri setleri ile temsil edilmektedir. Nota, müzikte sesin en küçük birimidir. Ölçüler ise belirli sayıda notayı gruplayan ve müziğin ritmini belirleyen yapılardır. Her ölçü, belirli bir vuruş sayısına sahiptir ve bu vuruşlar ile müzikal içerikteki zaman akışı düzenlenmektedir. Parça ifadesi ise birden fazla ölçüden oluşan, tamamlanmış bir müzikal eseri temsil etmektedir. Müzikal içeriklerde hiyerarşik ölçü-nota ilişkisi Şekil 2.4'te temsili bir müzikal ile desteklenerek sunulmaktadır.



Şekil 2.4: Hiyerarşik modelleme ile müzikal içeriğin gösterimi

Müzik üretiminde hiyerarşik gösterimin kullanımı üretici modellerin gelişimine yeni bir soluk getirmiştir. Dai ve ark. müzikal içeriği ölçü, cümle ve bölüm olmak üzere 3 seviyede inceleyen MusicFrameworks yöntemini önermiştir [21]. Yöntemin uzun-vadeli ilişkileri öğrenmede rakiplerine göre rekabetçi sonuçlar sunduğu ve birçok müzikal türde verimli çalışabildiği görülmüştür. Tez çalışmasında da bu hiyerarşik gösterimden esinlenilerek, müzikal ölçü seviyesindeki ilişkiler yüksek seviye müzikal değişkenler olarak belirlenmiş ve nota seviyesinde nota üretici modellerin uzun-vadeli ilişkileri güçlendirilmiştir.

2.2. Uzun-Vadeli Bağımlılıklar Problemi

Sıralı veriler, belirli bir zaman aralığında gözlemlenen ve sıralı bir şekilde depolanan değerler dizisi olarak ifade edilmektedir [22]. Bu veriler zamansal, uzamsal veya hiyerarşik bir sıraya sahip olabilmektedir. Literatürde hisse senedi fiyatları, hava durumu ölçümleri, sensör verileri veya müzik ve konuşma tanıma gibi çeşitli alanlarda sıralı veri gösterimi karşımıza çıkmaktadır [23]-[25]. Sıralı bağımlılıklar, bir sıralı veri setindeki ardışık değerler arasındaki ilişkilerdir. Bu ilişkilerin temsili dizi uzunluğu arttıkça zorlaşmakta ve uzun-vadeli bağımlılıklar problemi olarak adlandırılan problem ortaya çıkmaktadır [8].

Müzik uzun sıralı veriler arasındaki ilişkilere dayanan güçlü bir hiyerarşik yapıya sahiptir. Melodi, eşlik veya parçanın genel formu gibi kavramlar, müzikal içeriğin uzun vadeli ilişkilerinin güçlü temsiline bağlıdır [26]. Bir müzikal içerikte tekrar eden ve gelişen kısa melodik fikirler, çok daha sonra dönüştürülmüş veya armonize edilmiş olarak yeniden ortaya çıkabilmektedir [27]. Diğer yandan akor dizileri de uzun süre boyunca kalıpları takip edebilmekte ve bir beklenti ve çözüm duygusu yaratmaktadır.

Derin Öğrenme tekniklerinin gelişmesi ile son on yılda Yapay Zekâ ile üretilen müzikal içeriklerin genel kalitesinde iyileşme gözlemlenmektedir [7][26]. Bu durumun başlıca sebeplerinden biri, gelişen GPU kaynakları ile çok daha yüksek boyutlu verilerin işlenebilmesinin mümkün hale gelmesidir. Üretici yapay zekâ yöntemlerinin popülerleşmesi ile kodlayıcı-çözücü ağlar, yinelemeli sinir ağları ve dikkat ağları gibi birçok yöntem, yapay müzik üretimi probleminde de kullanılmaya başlamıştır.

Kodlayıcı-çözücü ağlar, kodlayıcı ve çözücü ağ adı verilen iki ağdan oluşmaktadır. Kodlayıcı ağlar girdi içeriğini bağlam vektörü adı verilen sabit uzunlukta yoğun bir gösterim ile kodlamaktadır. Çözücü ağlar ise bağlam vektörünü bir gürültü vektörü ile besleyerek başlangıçta sağlanan girdi vektörünü tekrar elde etmeye çalışmaktadır. Gürültü vektörü genellikle rastgele bir vektör olarak seçilirken, daha iyi gürültü vektörü seçimi yoluyla model öğrenimini iyileştirmeye yönelik çalışmaların da öğrenim başarımını artırdığı görülmektedir [28],[29]. Chung ve ark., kodlayıcı-çözücü mimarilerin özellikle müzikal verinin yoğun bir bağlam vektöründe temsiline olanak sağlaması ve verideki gürültünün azaltılmasındaki başarısını vurgulamaktadır [30]. Wu ve arkadaşları, hiyerarşik bir çözüm oluşturmak için iki farklı sinir ağı kombinasyonu kullanan bir yöntem önermektedir [31]. Yöntem ile CNN tabanlı Yapı Üretim Ağları kullanılarak müzik içeriğindeki tekrarlar kodlanmakta, daha sonra bu

değerler kullanılarak RNN tabanlı Melodi Üretim Ağı ile müzikal içerik üretimi gerçekleşmektedir. Bu modeller kısa-vadeli ilişkileri öğrenmekte başarılı olmakla birlikte, uzun-vadeli bağıntıları yakalamada yetersiz kaldıkları görülmektedir [29],[32].

Görüntü işleme problemlerinde yüksek başarım gösteren ve farklı alanlarda da yansımaları sıklıkla görülen konvolüsyonel sinir ağları (CNN) da müzikal içerik üretiminde literatürde değerlendirilen çözümlerden biridir. CNN'ler, aynı anda birçok işlemi yapabilme yetenekleri sayesinde büyük veri setlerinde bile hızlı ve verimli bir şekilde çalışabilmektedirler. Yöntemin MIR'da en önemli artısı müziğin ses dalgalarının görsel temsillerindeki yerel desenleri ve müzikteki zaman-frekans temsillerini modellemeki yüksek başarımıdır [33], [34]. Ancak, konvolüsyonel mimarilerin yerel bilgiye odaklanma problemi sebebiyle uzun-vadeli bağıntıları öğrenmede yetersiz kaldığı ve zamansal bağlamda uzak ilişkileri yakalayamadığı görülmektedir [7].

Yinelemeli sinir ağları (RNN) ve uzun-kısa süreli bellek (LSTM) ağları sıralı verilerde çalışmak üzere tasarlanmış mimariler olup, bu verilerdeki uzak ilişkileri daha yüksek başarımla temsil edebilmektedirler [27]. Böylece müziğin zaman içindeki dinamikleri etkili bir şekilde modellenebilmekte ve sınıflandırma, karar-destek sistem tasarımı veya müzikal içerik üretimi gibi problemlerde hem ritmik hem de melodik yapının matematiksel olarak daha etkin modellenebilmesi mümkün hale gelmektedir [35]. Müzikal içerik üretimi problemi özelinde ise yöntemin geleneksel sinir ağlarına göre çok daha güçlü bir verim sunduğu ve literatürde en çok tercih edilen model mimarilerinden olduğu görülmektedir [36]-[38]. Bununla birlikte RNN'ye dayalı modellerin de yeterince uzun zamansal diziler oluşturmakla görevlendirildiklerinde kaybolan gradyanlar problemi ile karşı karşıya kaldıkları görülmüştür [39]. Kaybolan gradyanlar problemi, derin sinir ağı geri yayılım algoritması ile ağırlıklarını güncellerken, sinirler arasındaki gradyan değerlerinin katmanlar boyunca küçülmesi ve sonuçta sıfıra yakınsamasından kaynaklanmaktadır. Bu durum da ağın daha derin katmanlarında öğrenme sürecini zorlaştırmakta ve modelin performansını olumsuz etkilemektedir [26],[27]. Daha da uzun LSTM ve Geçitli Tekrarlayan Birim (GRU) tabanlı çözümler, standart RNN'lere kıyasla sinir hücrelerinde ek bir bellek yapısı uygulayarak gradyanların kaybolması sorununa çözüm sunmakta ve uzun zaman serilerinde daha yüksek temsil başarımı ortaya koymaktadır [26],[27]. Liang ve ark. müzik verilerini müzikal ölçü seviyesinde müzikal özellik vektörüne dönüştürerek RNN'lerdeki bu sınırlamaları aşmayı hedeflemiş ve tek-enstrümanlı içerik üretebilen MIDI-Sandwich yöntemini önermiştir [40]. HCVAE ve HCGAN isimli iki alt

modelin birleşiminden oluşan MIDI-Sandwich, üretilen içerik uzunluğunu geliştirmeyi başarmıştır. Çalışmanın çok enstrümanlı versiyonu, MIDI-Sandwich2 olarak sunulmuş ve her biri bir başka enstrüman üzerinde eğitilen RNN ağlarının çok-modelli füzyon tekniği ile birleştirilmesini önermiştir [41]. Bu sayede uzun vadeli yapıdaki kazanımlar korunurken birden fazla enstrümana sahip parçaların incelenebileceği öne sürülmüştür. Boulanger-Lewandowski ve ark. da çok boyutlu sıralı verilerdeki uzun-vadeli bağımlılıkları modellemede RNN ve LSTM yöntemlerini kullanmıştır [36].

Son olarak, transformer modelleri, doğal dil işleme alanında büyük başarılar elde etmiş ve müzik üretiminde de kullanımı yaygınlaşmaya başlamıştır. Transformer modelleri, dikkat mekanizmaları sayesinde uzun-vadeli bağıntıları öğrenmede yüksek başarım sunmaktadır [42]. Dikkat mekanizmaları, verinin tamamı üzerinde paralel olarak çalışarak büyük veri setlerini hızlı bir şekilde işleyebilmekte ve uzun-vadeli ilişkileri öğrenme çok yüksek başarım sunmaktadırlar [42][43]. Bu durum, özellikle RNN tabanlı algoritmalara göre model eğitiminin daha hızlı gerçekleşebilmesine olanak sağlamaktadır. Ancak, transformer modelleri genellikle büyük veri setleri ve yüksek hesaplama gücü gerektirmektedir. Tez çalışmasında RNN, LSTM ve Transformer tabanlı dikkat ağları (Attention) yöntemleri müzikal içerik üretimi problemleri için geliştirilen yöntemlerde kullanılmıştır.

2.3. Yinelemeli Sinir Ağları

Yinelemeli Sinir Ağları, zaman serileri veya sıralı verilerle çalışmak için tasarlanmış özel bir sinir ağı türüdür. RNN'ler, önceki tüm zaman adımlarından gelen bilgiyi hatırlayabilme yetenekleri sayesinde, doğal dil işleme, MIR, konuşma tanıma ve zaman serisi tahminleri gibi çeşitli uygulamalarda yaygın olarak kullanılmaktadır. Bu tür ağlar, verilerin ardışık bağımlılıklarını ve zaman içindeki ilişkilerini modelleyebilme kabiliyetleri ile öne çıkmaktadırlar. RNN'ler, giriş verilerini adım adım işlerken, her bir adımda gizli durumları (hidden states) güncelleyerek bilgiyi hatırlamak ve yeni bilgiyle birleştirmek için tasarlanmıştır. RNN'in temel bileşeni, gizli durum (hidden state) ve bu durumun güncellenme şeklidir (2.1).

$$h_t = \sigma(W_{hx}X_t + W_{hh}h_{t-1} + b_h) \tag{2.1}$$

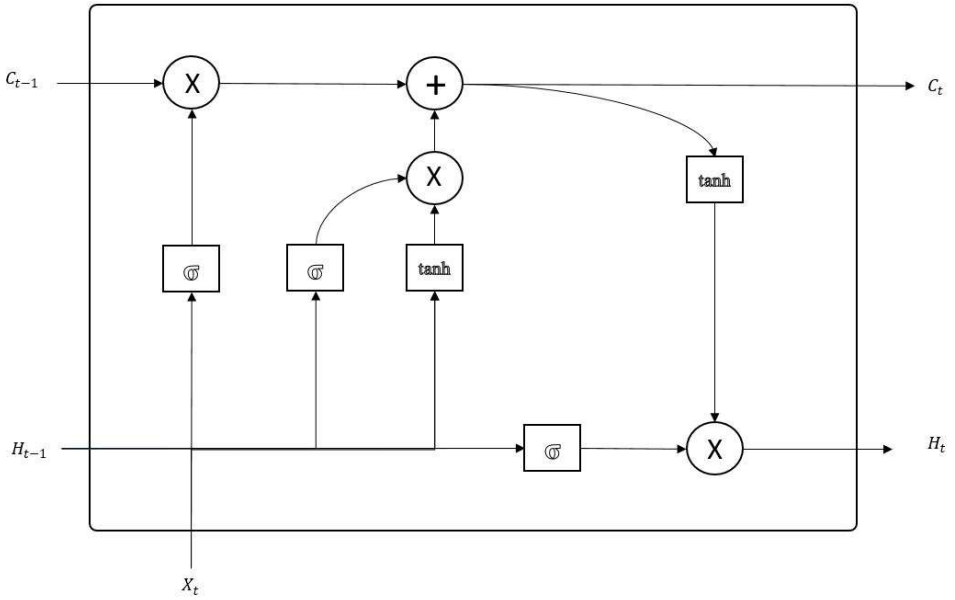
Burada h_t mevcut zaman adımındaki gizli durumu, x_t mevcut zaman adımındaki girdi vektörünü, W_{hx} girdi ağırlık matrisini, W_{hh} gizli durum ağırlık matrisini, b_h bias terimini ve σ aktivasyon fonksiyonunu temsil eder. Bu denklemin en önemli özelliği, h_{t-1} ile önceki gizli

durumun mevcut gizli duruma olan katkısını ifade etmesidir. Böylece, ağ geçmiş bilgiyi saklayabilmekte ve yeni bilgilerle birleştirerek öğrenme sürecini devam ettirebilmektedir.

RNN'ler sıralı verilerde verimli çalışmakla birlikte, sıralı veri uzunluğu arttıkça uzak veriler arasındaki ilişkilerin temsili zorlaşmaktadır. Bu sorun, "vanishing gradient" (kaybolan gradyan) ve/veya "exploding gradient" (patlayan gradyan) problemleri olarak bilinmektedir [27]. Eğitim süreci sırasında, gradyanlar çok küçük veya çok büyük hale gelebilmekte, bu da ağın uzun dizilimleri öğrenmesini zorlaştırmaktadır. Bu sorunları aşmak için LSTM (Long Short-Term Memory) ve GRU (Gated Recurrent Unit) gibi gelişmiş RNN türevleri geliştirilmiştir. Bu modeller, uzun vadeli bağımlılıkları daha etkili bir şekilde öğrenmek için RNN'lere ek araçlar eklemektedirler. Çalışma kapsamında LSTM yöntemi RNN ile birlikte temel yöntemler olarak kullanılmıştır.

2.4. Uzun-Kısa Dönem Hafıza Sinir Ağları

LSTM hücresinin temel bileşenleri giriş kapısı (input gate), unutma kapısı (forget gate) ve çıkış kapısıdır (output gate). Bu kapılar, hücrenin hangi bilgiyi saklayacağı, hangi bilgiyi güncelleyeceği ve hangi bilgiyi unutacağına karar vermektedir. LSTM hücresinin durumu C_t, hücrenin en temel bileşenidir ve zamansal içerik bilgisini saklamak için kullanılmaktadır. LSTM hücreleri, unutma kapısının sağladığı avantaj ile, eski bilgilerin değerlemesini her yeni bilgi varışında yenilemekte ve RNN'lere göre çok daha uzun sıralı verileri işleyebilmektedir. Standart bir LSTM hücresinin mimari tasarımı Şekil 2.5'te görülmektedir.



Şekil 2.5: Uzun-kısa dönem hafıza (LSTM) hücresi

Hücredeki ilk aşama, hangi bilgilerin unutulacağının kararının verilmesidir. Unutma kapısı, önceki gizli durumu (h_{t-1}) kullanarak unutma kararını vermektedir (2.2). Bu kapı, önceki gizli durum (h_{t-1}) ve mevcut girdi (x_t) bilgilerini kullanarak σ fonksiyonu ile 0-1 aralığına aktarılmaktadır. 0'a yakın değerler bilginin unutulması, 1'e yakın değerler ise unutulmaması gerektiğine işaret etmektedir.

$$f(t) = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{2.2}$$

Sonraki aşama, yeni bilginin hücre durumuna eklenme kararının verilmesi olup, bu aşamada giriş kapısı kullanılmaktadır (2.3). Giriş kapısı önceki gizli durum ve mevcut girdi bilgilerini kullanmakta ve σ fonksiyonu ile 0-1 aralığında sonuç üretmektedir. Yeni bilginin 0'a yakın değerler alması reddedileceği, 1'e yakın değerler alması ise kabul edilebileceğini göstermektedir.

$$i(t) = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2.3}$$

Yeni gizli durumun (h_t) hesaplanabilmesi için öncelikle hücre durum adayının ($\tilde{C}$) hesaplanması (2.4) ve hücre durumunun güncellemesi (2.5) gerekmektedir. Önceki gizli durum ve mevcut girdi, $tanh$ (hiperbolik tanjant) aktivasyon fonksiyonundan geçirilir. Elde edilen ($\tilde{C}$), hücre durumuna eklenecek yeni bilgilerin adayıdır. Hücre durumu ise, unutma kapısının (f_t)

çıktısı ile giriş kapısının çıktısı kullanılarak güncellenir. Bu adımda, eski hücre durumu ile yeni hücre durumu adayı, birlikte yeni hücre durumunu oluşturmaktadır.

$$\tilde{C}(t) = tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{2.4}$$

$$C(t) = \left(f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t\right) \tag{2.5}$$

Çıkış kapısı, hücre durumunun hangi kısmının çıkış olarak kullanılacağını belirlemektedir. Çıkış kapısında da önceki gizli durum ve mevcut girdi bilgileri kullanılmakta ve sonuç bir sigmoid fonksiyonundan geçirilmektedir (2.6).

$$o(t) = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{2.6}$$

Son olarak, güncellenmiş hücre durumu ve çıkış kapısının çıktısı kullanılarak yeni gizli durum hesaplanmaktadır. Hücre durumu, $tanh$ aktivasyon fonksiyonundan geçirilmekte ve çıkış kapısının çıktısı ile çarpılarak gizli durum güncellenmektedir (2.7).

$$h(t) = o_t \cdot \tanh(C_t) \tag{2.7}$$

2.5. Öz-dikkat Ağları (Self-Attention)

Self-attention (öz-dikkat) mimarisi, Transformer modern sinir ağı mimarilerinin temelini oluşturan yapıdır ve doğal dil işleme, makine çevirisi, metin oluşturma gibi birçok alanda uzun-vadeli bağıntıları yakalamada devrim yaratmıştır [42]. Öz-dikkat, sıralı verilerde bir girdi öğesinin diğer tüm girdi öğeleriyle olan ilişkisini dikkate almakta ve bu ilişkileri ağırlıklı ortalamalar kullanarak hesaplamaktadır. Öz-dikkat mekanizması, Query (sorgu), Key (anahtar) ve Value (değer) olmak üzere üç temel bileşenden oluşmaktadır. Hesaplamalar sırasında her bir girdi vektörü (X_i) bu üç bileşen vektöründe temsil edilmek üzere dönüştürülmektedir. Dönüşüm denklemleri (2.8), (2.9) ve (2.10)'da sunulmuştur. Burada W_Q, W_K ve W_V öğrenilebilir ağırlık matrislerini ifade etmektedir.

$$Q = X \cdot W_Q \tag{2.8}$$

$$K = X \cdot W_K \tag{2.9}$$

$$V = X \cdot W_V \tag{2.10}$$

Öz-dikkat mekanizmasının ana adımı, her bir Q_i sorgu vektörü ile tüm K_j anahtar vektörleri arasındaki benzerlikleri hesaplamaktır. Bu benzerlikler, nokta çarpımı (dot product) ve ardından softmax fonksiyonu kullanılarak hesaplanmaktadır. Bu işlem, her bir sorgunun diğer tüm anahtarlarla olan ilişkilerini belirlemektedir (2.11). Denklemde sorgu ve anahtar matrislerinin nokta çarpımı (QK^T) ile, anahtar vektörünün boyutu (d_k) ile temsil edilmektedir.

$$Dikkat\ skoru(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2.11}$$

Öz-dikkat mekanizmasının çıktısı, her bir sorgu vektörünün diğer tüm değer vektörleri (V) ile ağırlıklı ortalamasıdır (2.12). Denklem (2.11)'de hesaplanan dikkat ağırlıkları, öz-dikkat değer matrisi (V) ile çarpılarak içerik vektörü elde edilmektedir. Böylece yöntem her bir giriş öğesinin diğer tüm giriş öğeleriyle ilişkilerini tek bir gösterimde dikkate alabilmektedir. Bu sayede, birbirinden ne kadar uzakta olursa olsun, her bir girdi arasında doğrudan bir ilişki kurulmuş olmaktadır.

$$\text{İçerik vektörü} = Dikkat\ skoru * V \tag{2.12}$$

Öz-dikkat ağlarının en büyük avantajlarından biri, paralel işlenmeye müsait matris gösterimini kullanıyor olmalarıdır. Geleneksel RNN'ler ve LSTM'lerin aksine, self-attention mekanizması tüm giriş öğelerini aynı anda işleyebilmekte ve özellikle büyük veri setleri ve uzun dizilerle çalışırken büyük bir hız ve verimlilik sağlamaktadır. Bununla birlikte, öz-dikkat vektörlerinin bellek kullanımı, girdi dizisinin uzunluğu N için $O(N^2)$ karmaşıklığa sahiptir. Bu durum, uzun dizilerle çalışırken bellek yetersizliğine yol açabilmekte veya dizi uzunluğunun donanımsal kısıtlar ölçüsünde sınırlandırılmasını gerektirebilmektedir. Ayrıca, büyük boyutlu ağırlık matrisleri ve karmaşık hesaplamalar, eğitim sürecini de uzatabilmektedir.

2.6. Çok-başlıklı Dikkat Ağları (Multi-Head Attention)

Öz-dikkat mekanizması, sinir ağlarının sıralı verilerle çalışırken her bir öğenin diğer tüm öğelerle olan ilişkilerini dikkate almasını sağlar. Ancak, model farklı ilişkileri ve desenleri paralel olarak öğrenmede yetersiz kalabilmektedir. Tek bir dikkat başlığı, her bir öğe arasındaki ilişkileri belirlemek için tek bir dikkat mekanizması kullanmakta ve bu, karmaşık ve çok

boyutlu ilişkileri tam anlamıyla modelleyememektedir. Bu sınırlamaların üstesinden gelmek için çok-başlıklı dikkat (multi-head attention) mekanizması önerilmiştir [42].

Çok-başlıklı dikkat mekanizması, öz-dikkat mekanizmasının bir genişlemesidir ve modelin farklı dikkat başlıklarını (attention heads) kullanarak farklı ilişki türlerini paralel olarak öğrenmesine olanak tanımaktadır. Her bir dikkat başlığı, giriş verilerini farklı bir alt uzaya projekte etmekte ve bu alt uzayda kendi dikkat hesaplamalarını yapmaktadır (2.13). Bu durum, modelin özellikle çok değişkenli bilgi uzaylarında farklı bilgi türlerini ve ilişkileri daha iyi yakalamasını sağlamaktadır. Çok-başlıklı dikkat mekanizmasında, her bir başlığın kendi sorgu (query), anahtar (key) ve değer (value) matrisi bulunmaktadır. Bu matrisler, girdi vektörlerini farklı projeksiyonlara dönüştürmektedir. Böylece her biri kendi dikkat skorlarını hesaplayan başlıklar kullanılarak girdi vektörleri farklı perspektiflerden incelenebilmektedir (2.14).

$$Q_i = X \cdot W_{Q_i}, K_i = X \cdot W_{K_i}, V_i = X \cdot W_{V_i} \tag{2.13}$$

$$Dikkat\ skoru_i(Q_i, K_i, V_i) = softmax\left(\frac{Q_i {K_i}^T}{\sqrt{d_k}}\right) V_i \tag{2.14}$$

Çok başlıklı dikkat ağlarında bulunan başlıkların her birinin çıktısı nihai bir aşamada birleştirilmekte ve son bir lineer dönüşüm uygulanmaktadır (2.15). Burada W_o, çıktı ağırlık matrisini, Birleştir (concat) ise başlıkların birleştirilmesi işlemini ifade etmektedir. Başlık çıktıları ise V_i değerlerinin ağırlıklı bir toplamı olarak hesaplanmaktadır. Hesaplamada kullanılan ağırlıklar ise, ilgili başlığın K_i anahtarlarının Q_i sorgusu ile yakınlığına (relevance) göre belirlenir (2.11).

$$MTHA(Q, K, V) = Birleştir(başlık_1, \dots, başlık_h) W_o \tag{2.15}$$

Çok-başlıklı dikkat mekanizmasının en önemli avantajı, modelin farklı başlıklar kullanarak çeşitli ilişki türlerini ve desenleri paralel olarak öğrenebilmesidir. Bu, modelin daha zengin ve esnek temsil öğrenimleri elde etmesini sağlamaktadır. Ayrıca, her bir başlığın farklı bir alt uzayda çalışması, modelin bilgi çeşitliliğini artırmakta ve genelleme yeteneğini geliştirmektedir.

3. YÖNTEM

MIR alanında yapılan çalışmaların büyük çoğunluğu, sembolik sıralı veri gösterimini kullanmakta ve zamana dayalı değişkenlere bağlı analizler sunmaktadır. Müziğin karakteristik bir özelliği ise hiyerarşik ve çok-katmanlı bir yapıya sahip olmasıdır. Notalar, ölçüler, bölümler ve hareketlerden oluşan farklı düzeylerde incelenebilen müzikal içerikler, düzey içi ve düzeyler arası ilişkilerin belirlenmesi ile daha kapsayıcı bir temsil ile incelenebilmekte ve müziğin altında yatan matematiği anlamamızda yeni fırsatlar oluşturmaktadır.

Tez çalışmasında tek-sesli melodi üretimi problemine yönelik iki yöntem geliştirilmiştir. Öncelikle MTHA-LSTM yönteminde sıralı gösterimler kullanılarak nota seviyesinde melodik içerik üretimi amaçlanmıştır. Böylece uzun-vadeli bağımlılıkların daha etkili temsil edilmesi sağlanmış ve uzun içeriklerde dahi değişkenler arasındaki bağlamın korunabilmiştir. İkinci yöntemde ise nota üretimi problemi, müzikal bilgi çıkarımı ile desteklenerek hiyerarşik ve çok-katmanlı GAMA-DL yöntemi geliştirilmiştir. GAMA-DL yöntemi, müzikal ölçü ve nota seviyesinde entegre çalışan üç DÖ yönteminin kullanıldığı genel mimari ismi olup, mimarideki alt-problemlerin her birinde içerik üretimi amacıyla MTHA-LSTM yöntemi kullanılmıştır.

MTHA-LSTM ve GAMA-DL yöntemlerinin ürettiği tek-sesli melodileri çok-sesli müzikal içeriğe dönüştürebilmek için ise eşlik üretici modül geliştirilmiştir. Çalışmada kullanılan veri setleri Bölüm 3.1'de, MTHA-LSTM Bölüm 3.2'de, GAMA-DL Bölüm 3.3'de, eşlik üretici modül ise Bölüm 3.4'te sunulmaktadır.

3.1. Veri Setleri ve Ön İşleme

Çalışmadaki analizler, iki farklı klasik tonal batı müziği veri seti kullanılarak gerçekleştirilmiştir. İlk veri kümesi, Klasik Piyano Midi (KPM) olarak adlandırılmakta olup, 25 dünya çapında ünlü klasik müzik bestecisinin MIDI formatındaki piyano kompozisyonlarını içermektedir ve CC-by-SA lisansı ile paylaşılmaktadır [44]. İkinci veri kümesi ise Google Magenta ekibi tarafından işlenerek kamuya açılan MAESTRO veri kümesidir [45]. MAESTRO veri kümesi toplamda yaklaşık 200 saatlik piyano performans kaydının MIDI formatında işlenmiş sürümlerini içermektedir.

Her iki veri kümesi de açık veri kümesi sınıfında değerlendirilmekte ve Web ortamında araştırmacıların kullanımına sunulmaktadır. MIDI yapıda kodlanmış müzikal bilginin derin öğrenme modellerinde kullanılabilecek matematiksel değişkenlere dönüştürülmesinde "music21" müzik işleme kütüphanesi kullanılmıştır [46].

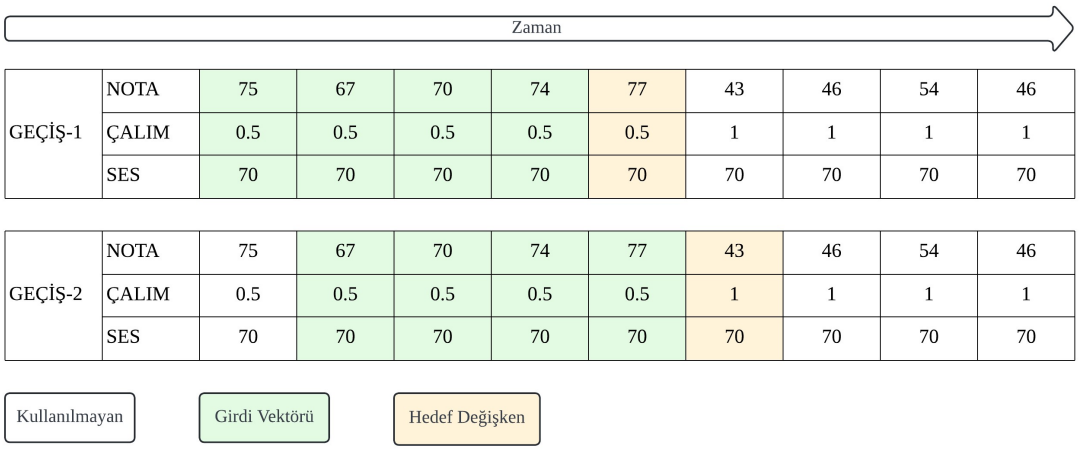
MIDI müzikal eserlerden ham veri elde edilmesinde nota seviyesinde üç değişken (nota, çalım süresi, ses seviyesi) kullanılmıştır. Kategorik bir değişken ile temsil edilen nota, çalınan sesin frekans sınıfını belirtmektedir. Çalışmada piyano enstrümanında çalınabilen ve MIDI ortamında karşılığı bulunan 88 farklı nota frekans sınıfı kullanılmıştır. Parçalardaki akor yapıları, temel notaları kullanılarak tek notaya indirgenmiştir. Veri kümesinin ikinci değişkeni olan çalım süresi, notanın çalım süresini ifade etmekte ve sayısal bir değişkenle temsil edilmektedir. Ses seviyesi değişkeni ise notanın çalım sırasındaki ses seviyesini ifade etmekte ve yine sayısal bir değişken ile temsil edilmektedir. Ham hali MIDI kodlanmış müzikal içeriklerden oluşan KPM ve MAESTRO veri setleri için elde edilen veri kümesi özellikleri Tablo 3.1'de sunulmaktadır. 20 kata yakın daha fazla özgün girdi satırı içeren MAESTRO veri seti, KPM'ye göre çok daha büyük bir müzikal bilgi uzayında analiz yapma imkânı sunmaktadır. Özgün nota sayısı bakımından KPM'de daha az özgün notanın bulunduğu görülmekte olup, ses zenginliği açısından da MAESTRO'da daha zengin bir içerik bulunduğu görülmektedir.

Tablo 3.1: KPM ve MAESTRO veri setlerinin istatistiki özellikleri

Veri Seti	KPM	MAESTRO
Parça Sayısı	295	1276
Özgün Nota Sayısı	83	88
Girdi vektör büyüklüğü	381248	7040164

Girdi vektörleri sıralı vektörlerden oluşmaktadır. Ham verisetinden sıralı vektör elde etmek amacıyla kayar pencere yöntemi kullanılmıştır. Kayar pencere yöntemi, belirli bir p uzunluğunda pencere oluşturup, bu pencereyi sıralı girdi vektörü boyunca kaydırarak her adımda verinin bir alt kümesini elde etmektedir. Pencere boyutu, tahmin edilecek bağımlı değişkeni elde etmede dikkate alınacak önceki ardışık değerlerin sayısıdır. Pencere kaydırılması, belirli bir adım boyutu t büyüklüğünce gerçekleştirilmektedir. Her geçişte, en eski veri noktası girdi vektöründen çıkarılmakta ve pencerenin sonuna yeni bir veri noktası eklenmektedir. Adım boyutu tez çalışması kapsamında bir olarak belirlenmiştir. Şekil 3.1, kayar

pencere yöntemi ile ham verisetinden girdi vektörü ve hedef değişken elde edilişini görsel olarak sunmaktadır. Her geçişte, p uzunluğunca ardışık adım girdi vektörünü, sonraki adım ise hedef vektörünü oluşturmaktadır.



Şekil 3.1: Kayar pencere yöntemi ile girdi vektörü - hedef değişken eldesi

3.2. MTHA-LSTM

MTHA-LSTM yönteminde ilk aşamada LSTM katmanı ile bir içerik vektörü elde edilmekte, daha sonra ise bu vektör MTHA katmanı ile ağırlıklandırılmaktadır. Böylece uzun-vadeli bağımlılıklar güçlü bir şekilde temsil edilebilirken, aynı zamanda anlamlı ve akıcı yapay müzikal içerikler üretilebilmektedir. Müzikal içerik üretici model, sembolik gösterim ile çıkarılan öznitelikleri kullanılmaktadır. Bölüm 3.2.1, yöntemde kullanılan veri gösterimini açıklamakta, Bölüm 3.2.2 ise yöntemin mimari tasarımını sunmaktadır.

3.2.1. Veri Gösterimi

Nota değişkeni, piyano ile çalımı gerçekleştirilebilen 88 ses frekansı ve sus olmak üzere toplamda 89 farklı değer alabilmektedir. Parçalarda görülme sıklığı incelendiğinde, nadir kullanılan notalar bulunduğu görülmüştür. Tüm müzikal bilgi uzayında toplamda 100'den daha az karşılaşılan notalar "nadir notalar" olarak belirlenmiş ve bu notalar eğitim verilerinden çıkarılmıştır. Böylece, YZ modellerinin öğrenme başarımı ile nihai içeriğin müzikal kalitesine en az etkide bulunularak hesaplama karmaşıklığının azaltılması amaçlanmıştır. Çalım süresi ve ses seviyesi, nümerik değişkenler olduğundan, bu değişkenlerde normalizasyon uygulanmıştır. Veri normalizasyonu, veri setindeki değişkenlerin belirli bir aralıkta yeniden ölçeklendirilmesi işlemidir. DÖ algoritmaları, farklı ölçeklerde girdi değişkenleri ile beslendiğinde, değişkenlerin

birbirlerine baskınlık oluşturması ve model başarımının olumsuz etkilenmesi gözlemlenebilmektedir. Normalizasyon tekniği ile bu baskınlık durumu ortadan kaldırılarak tüm değişkenlerin model öğrenimine etkisi dengelenmekte ve belirli değişkenlere aşırı odaklanma durumu ortadan kaldırılmaktadır.

Min-Max normalizasyonu, verileri belirli bir aralığa yeniden ölçeklendirmek için kullanılan bir yöntemdir. Bu yöntem, her bir veri noktasının kendi minimum ve maksimum değerleri arasında normalize edilmesi prensibine dayanmaktadır (3.1). Çalışma kapsamında nümerik değişkenler çalım süresi ve ses seviyesi için normalizasyon işlemleri, değişkenler [0-1] aralığına aktarılacak şekilde gerçekleştirilmiştir.

$$\acute{X} = \frac{X - X_{min}}{X_{maks} + X_{min}} \tag{3.1}$$

3.2.2. Mimari Tasarım

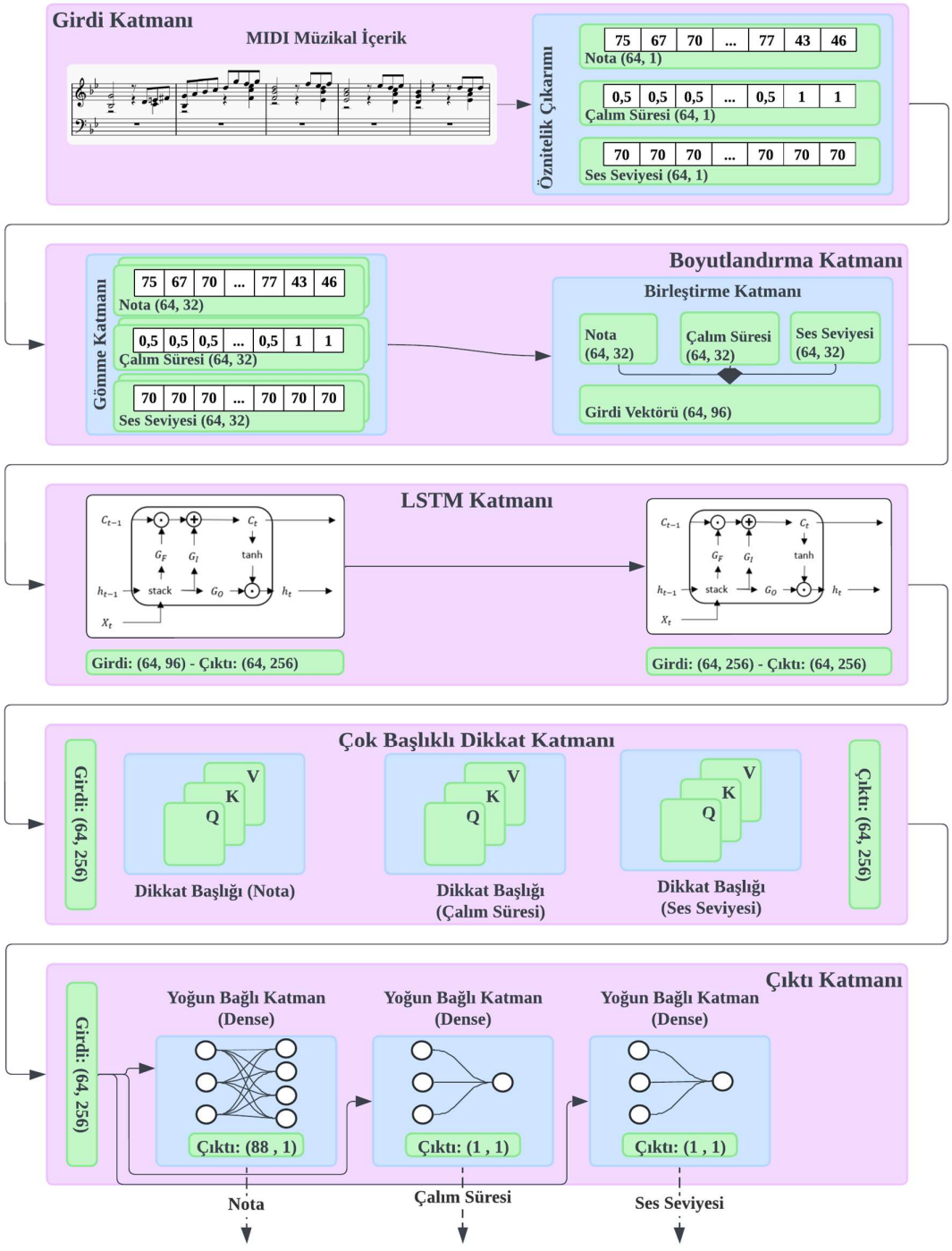
MTHA-LSTM, melodi üretme problemini iki katmanlı bir mimariyle ele almaktadır. Bu yöntemde öncelikle LSTM katmanları ile bir içerik vektörü elde edilmekte, daha sonra ise MTHA katmanı ile içerik vektörü ağırlıklandırılmaktadır. Yapay Zekâ araçları ile içerik üretiminde otoregresif yöntemler ve otoregresif-olmayan yöntemler olmak üzere iki popüler strateji bulunmaktadır [47]. Otoregresif yöntemler, sıralı bir sekans üretiminde bir adımda önceki değerlere göre şartlanmış tek yeni değer üretmektedir. Böylece üretilen sekanslarda doğruluk seviyesi yüksek kalabilmekte ve gerçek bilgi uzayına yakın kesinlikte sonuçlara ulaşılabilmektedir. Konuşma dilleri arasında çeviri veya metin ve müzikal içerik üretme gibi problemlerde sıklıkla başvurulan yöntemin dezavantajı ise, geliştirilen modellerin sekans içeriğindeki her bir veri noktası için ayrı ayrı çalıştırılması gerekliliğidir. Her veri noktasının sonucu daha önceki veri noktalarının sonuçlarına bağlı olacağından, bu işlem paralelleştirmeye de müsait değildir. Bu sebeple toplam içerik üretim süresi otoregresif olmayan yaklaşımlara göre daha uzun olmaktadır.

Otoregresif olmayan modeller ise, tüm veri noktalarını aynı anda tek adımda veya birkaç adımda ele almayı hedeflemektedir [27][42][48]. Bu yaklaşım, aynı uzunlukta içerik üretimi için gerek duyulan toplam süreyi otoregresif yaklaşım ile geliştirilen bir yönteme göre önemli ölçüde azaltmaktadır. Bununla birlikte, veri noktaları arasındaki sıralı bağımlılıkların doğrudan bir matematiksel ilişkisi temsil aracılığıyla şartlanmadığından, üretilen içeriğin doğruluğu ve

gerçek bilgi uzayına yakınlığı düşük seviyede kalmaktadır. Tez çalışması kapsamında, müzikal kaliteyi artırarak sıralı verilerde uzun-vadeli ilişkilerin temsilinin iyileştirilmesi amaçlandığından, müzikal içerik üretimi problemi otoregresif yaklaşım ile ele alınmıştır. Bu amaçla geliştirilen MTHA-LSTM yönteminin mimari tasarımı Şekil 3.2'de sunulmaktadır.

Yöntem girdi katmanında MIDI ile kodlanmış müzikal içeriklerden "nota", "çalım süresi" ve "ses seviyesi" olmak üzere üç girdi vektörü elde etmektedir. Bu girdi vektörlerinin tamamı sıralı değişken tipine sahip olup (64, 1) boyutlarında vektörler ile temsil edilmektedir. Kategorik değişkenlerin işlenmesinde One-Hot Kodlama (One-Hot Encoding) ve Etiket Kodlama en popüler yöntemler olarak öne çıkmaktadır. One-Hot kodlamada her kategori, yalnızca biri "1" ve diğer kategoriler "0" değerini alacak şekilde ikili vektörler ile temsil edilmektedir. Bu durumda birçok sıfır değer içeren, seyrek matrisler elde edilmektedir. Her ne kadar kategoriler arasında baskınlık ve bir kategoriyi öne çıkarma durumlarına karşı etkin bir çözüm olsa da, yüksek sayıda kategori ile çalışıldığında hafıza karmaşıklığında olumsuz etkiye neden olabilmektedir.

Etiket kodlama, One-Hot kodlamaya bir alternatif olarak kategorik değişkenlerin gösteriminde kullanılmaktadır. Bu yöntemde her kategori benzersiz bir tamsayı ile temsil edilmekte ve bu sayede hafıza çok daha verimli kullanılmaktadır. Ayrıca bu şekilde kodlanan vektörler, gömme katmanlarına girdi olarak verilebilmektedirler. Gömme katmanları, kategorik değişkenleri daha düşük boyutlu yoğun vektörlere dönüştürmek amacıyla kullanılmaktadır.



Şekil 3.2: MTHA-LSTM Mimari Tasarım

MTHA-LSTM yönteminde etiket kodlama yöntemi kullanılmış ve elde edilen vektörler YZ modelinde gömme katmanları ile yoğun vektörlere dönüştürülmüştür. Gömme katman boyutu 32 olarak belirlenmiştir. Gömme katmanı, girdi boyutunu daha küçük vektörlere aktarmakta ve

farklı boyutlarda girdi vektörlerinin sabit uzunluklarda yapılandırılmasında kullanılmaktadır. Gömme aşaması sonucunda her bir sıralı değişken vektörü için (64, 32) ebatlarında boyutlandırılmış girdi vektörleri elde edilmektedir. Birleştirme katmanında ise bu vektörler birleştirilerek (64, 96) ebatında bir girdi vektörüne dönüştürülmektedir.

Girdi vektörleri arka arkaya iki LSTM katmanında işlenmekte ve içerik vektörü elde edilmektedir. Her iki LSTM katmanı da 256 nörona sahiptir. İlk LSTM katmanında (64, 256) boyutunda bir içerik vektörü elde edilmektedir. Bu vektör, LSTM yönteminin tüm sıralı çıktılarını içeren bir vektördür. İkinci LSTM katmanında ise yine (64, 256) ebatında vektör elde edilmektedir. İlk LSTM katmanı, giriş dizisindeki temel kalıpları ve ilişkileri yakalaması amacıyla kullanılmaktadır. Ancak, bu kalıplar nispeten basit veya kısa-uzunlukta ilişkileri içermektedir. İkinci LSTM katmanı, ilk katmandaki içerik vektörünü girdi olarak alarak girdi vektörlerinin öğrenilmiş gösterimlerindeki daha karmaşık veya daha yüksek seviyeli zamansal bağımlılıkları öğrenmeyi amaçlamaktadır. LSTM katmanlarının her ikisinde de nöron sayıları, sabit ve girdi vektör uzunluğunun iki katından fazla olacak şekilde ayarlanmıştır. Müzikal içerikler, hiyerarşik ve çok-seviyeli ilişkilerin oldukça güçlü olduğu yapılar olduğundan, birden fazla LSTM katmanı kullanımı, hiyerarşik ilişkilerin yakalanmasında avantaj sağlamaktadır. Her iki LSTM katmanında kullanılan nöron sayılarının belirlenmesinde deneysel sonuçlar, model karmaşıklığı ve kayar pencere uzunluğu etkili olmuştur. Nöron sayıları, kayar pencere uzunluğunun en az iki katı olacak şekilde hiperparametre optimizasyonuna tabi tutulmuştur. Fazla öğrenmenin önüne geçmek amacıyla her iki LSTM katmanından hemen sonra (drop-out) işlemi uygulanmaktadır. Drop-out işlemi, düzenlileştirme “regularization” tekniklerinden biri olup, DÖ modellerinin genelleştirme yeteneğini artırmak ve fazla-öğrenme durumunu önlemek için kullanılmaktadır [49]. Modeldeki belirli nöronlar drop-out işlemi ile her iterasyonda rasgele devre dışı bırakılmakta ve yöntemin nöronlara aşırı bağlı olmasının önüne geçilmektedir.

LSTM katmanlarında elde edilen içerik vektörü MTHA katmanına iletilmekte ve burada işlenerek ağırlıklandırılmaktadır. MTHA katmanına giriş ve çıkış yapan vektör boyutları aynıdır. Çıkış katmanında MTHA katmanında ağırlıklandırılmış olan içerik vektörü kullanılmakta ve her bir değişken için ayrı bir yoğun bağlantılı katman aracılığıyla çıkış değerleri elde edilmektedir. Farklı yoğun bağlantılı katman kullanımı ile değişkenlere özgü özelliklerin daha iyi öğrenilmesi sağlanmaktadır.

3.3. GAMA-DL

MTHA-LSTM yöntemi, uzun-vadeli ilişkileri yakalamada oldukça başarılı olsa da yöntemin gerek duyduğu yüksek hafıza karmaşıklığı sebebiyle ölçeklenebilirliği zayıf kalmaktadır. GAMA-DL ile MTHA-LSTM'in ölçeklenebilirliğini geliştirmek amacıyla hiyerarşik bir mimari önerilmektedir. GAMA-DL, yüksek seviyeli müzikal ölçü üretimi ve düşük seviyeli nota üretimi olmak üzere iki aşamaya sahip derin öğrenme yönteminin genel adıdır. Müzikal ölçüler, bir müzikal içeriğin zamansal bölümlere ayrılmasını sağlamaktadır [54]. Her bir müzikal ölçü bir dizi vuruş (beat) içermektedir. Müzikal ölçüler, müziğin ritmini ve yapısını düzenleyerek müzisyenlerin ve dinleyicilerin parçayı daha kolay takip etmelerini sağlamak amacıyla kullanılmaktadır.

Müzikal ölçü üretici ağ, GAMA-DL'in ilk aşaması olup üretilecek melodinin müzikal ölçüleri arasındaki ilişkileri belirlemektedir. Bu ağın çıktıları nota seviyesindeki üretici yöntemlerde kullanılmaktadır. Nihai melodi üretimi, ikinci aşamada çalışan nota üretici yöntemlerde gerçekleşmektedir. Her bir müzikal ölçü için sırayla ilişki durumuna bakılmakta, ilişki var ise müzikal ölçü koşullu nota üretici ağ, yok ise koşulsuz nota üretici ağ kullanılmaktadır. Koşulsuz nota üretici ağın görevi, geçmişteki notalara göre sonraki notayı tahmin etmektir. Bu kapsamda baz yöntemlere göre başarımı kanıtlanmış ve tez çalışmasının ilk yöntemi olarak önerilmiş olan MTHA-LSTM kullanılmıştır. Müzikal ölçü koşullu nota üretici ağın farkı, zamansal ilişkilere ek olarak müzikal ölçü ilişkilerini de dikkate almasıdır.

3.3.1. Veri Gösterimi

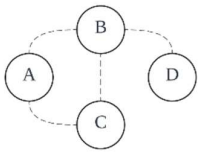
GAMA-DL yöntemi müzikal ölçü üretiminde çizge gösterimi, koşullu ve koşulsuz nota üretiminde ise sıralı gösterim kullanmaktadır. GAMA-DL koşulsuz nota üretici modülde sıralı veri gösterimi kapsamında, Bölüm 3.2.1'de sunulmuş olan MTHA-LSTM yönteminin veri gösterimi kullanılmaktadır. Müzikal ölçü üretici modülde kullanılan çizge gösterimi Bölüm 3.3.1.1'de, koşullu nota üretici modül veri gösterimi ise Bölüm 3.3.1.2'de sunulmaktadır.

3.3.1.1. Müzikal Ölçü Seviye Gösterim (Çizge)

GAMA-DL, müzikal ölçü seviyesinde çizge gösterimini kullanmaktadır. Bu sayede müzikal ölçü üretimi problemi, çizge üretimi problemine dönüşmektedir. Çizgeler, düğümler ve kenarlardan oluşan, karmaşık ilişkilerin temsilinde kullanılan veri yapılarıdır. Müzikal ölçü veri setinde her bir müzikal ölçü bir düğümü, ölçüler arasındaki ilişkiler ise kenarları ifade etmektedir. İlişki türü, bir kenar özniteliği olarak belirlenmiştir. Çizgelerin matematiksel olarak

gösteriminde kullanılan en popüler yöntemler komşuluk matrisi ve komşuluk listesidir (K). Komşuluk matrisleri, tüm düğümlerin diğer tüm düğümlere olan bağlantılarını gösteren iki boyutlu kare matrislerdir. İki düğüm arasında bağlantılılık durumu $O(1)$ zamanda çözülebilmektedir. Hafıza karmaşıklığı ise N düğümlü bir çizge için $O(N^2)$ olup, seyrek ilişkiye sahip çizgelerde verimsiz bir gösterim olmasına sebep olmaktadır. Komşuluk listelerinde ise her düğüm sadece komşu düğümlerini (bağlı olduğu düğümler) içeren bir liste oluşturulmaktadır. Bu sayede yoğun bağlantılı olmayan çizgelerde daha verimli bir hafıza karmaşıklığı sunmakta, bağlantılılık durumu ise düğüm sayısı d olacak şekilde $O(d)$ karmaşıklıkla belirlenebilmektedir. Müzikal ölçüler, yoğun bağlantılı çizgeler oluşturmadığından, çalışmada komşuluk listesi gösterimi tercih edilmiştir (Şekil 3.3).

(a) (b) (c)



Düğümler	A	B	C	D
A	0	1	1	0
B	1	0	1	1
C	1	1	0	0
D	0	1	0	0

Düğüm	Komşular
A	['B', 'C']
B	['A', 'C', 'D']
C	['A', 'B']
D	['B']

Şekil 3.3: Çizgelerde komşulukların gösterimi, (a) Çizge örneği, (b) Komşuluk matrisi gösterimi, (c) Komşuluk listesi gösterimi

Müzikal ölçü seviyesinde ilişkilerin tespiti yakın tarihli çalışmalarda çalışılmaya başlanan, popüler bir araştırma alandır [50][51]. Tez çalışmasında, literatürde önerilmiş müzikal ölçü ilişkileri arasından batı klasik tonal müzik alanında yüksek öneme sahip altı ilişki analiz edilmiştir. Bu müzikal ölçü ilişkileri, şöyle ifade edilmektedir:

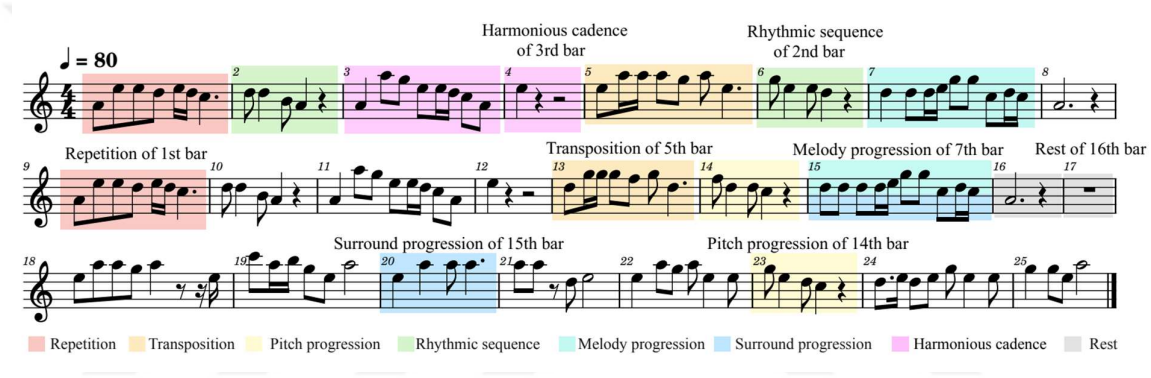
- Tekrar (Repetition): İki müzikal ölçünün ritmik, perde ve oktav açılarından birebir aynı olma durumu.

- Transpozisyon (Transposition): İki müzikal ölçünün ritmik olarak aynı ve nota dizileri arasındaki oktav farkının aynı olma durumu.

- Perde İlerlemesi (Pitch Progression): İki müzikal ölçünün ritmik olarak aynı ve nota perdelerinin en az yarısının aynı olma durumu. Nota oktavları farklı olabilir.

- Ritmik Sekans (Rhythmic Sequence): İki müzikal ölçünün ritmik olarak aynı ve nota perdelerinin yarısından azının aynı olma durumu. Nota oktavları farklı olabilir.
- Melodi İlerlemesi (Melody Progression): İki müzikal ölçü arasında en az 3 adet arka arkaya aynı perde ve ritimde nota bulunması.
- Sus (Silence): İki sessiz müzikal ölçü arasındaki ilişki.

Şekil 3.4, bir müzikal içerikte belirtilen ilişkileri görsel olarak sunmaktadır.



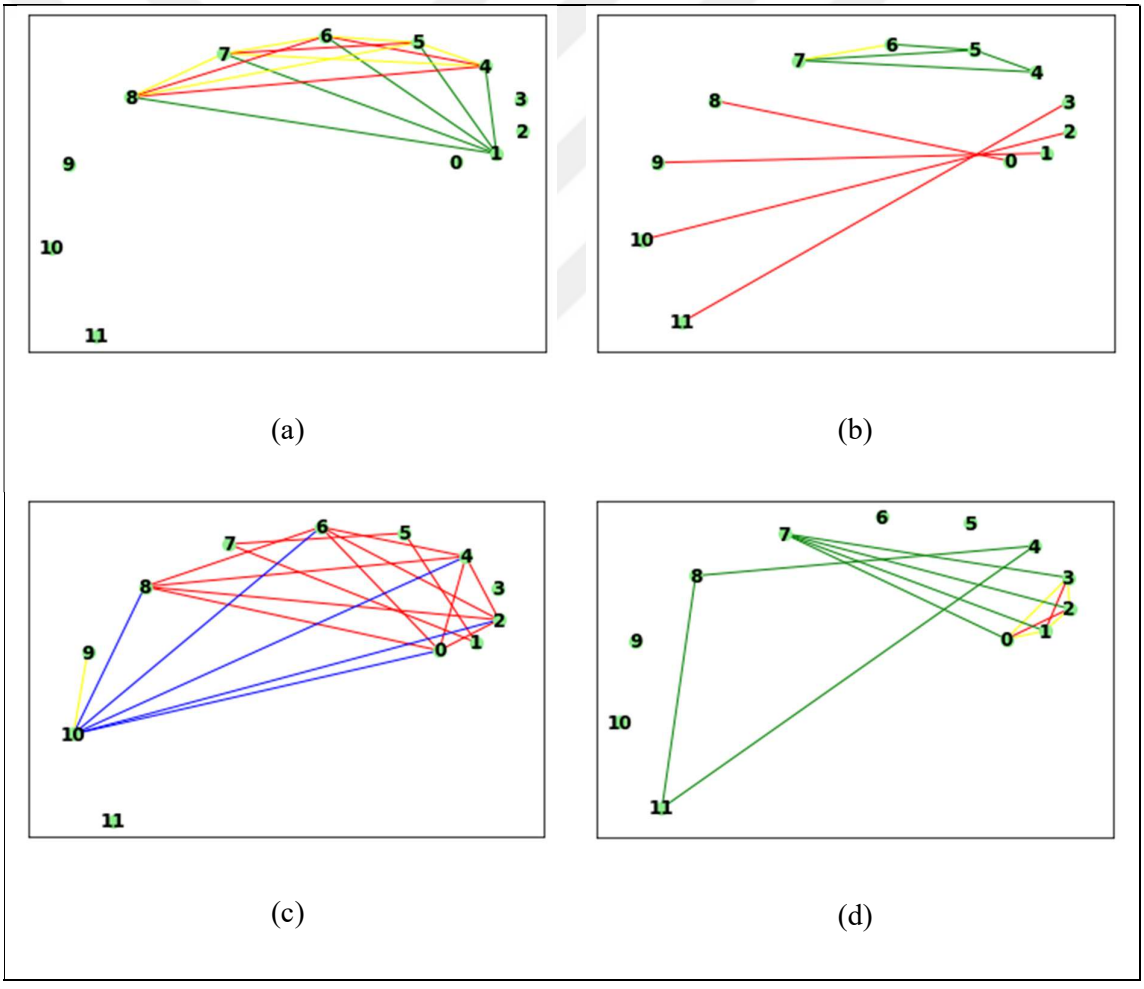
Şekil 3.4: Sembolik Gösterim ile Müzikal Ölçü İlişkileri [51]

Müzikal ölçü veri seti, parçaların on iki müzikal ölçülük alt kümelerini elde etmekte ve bu büyüklük içerisinde müzikal ölçü ilişkilerini belirleyerek çizge veri yapısında temsil etmektedir. Literatürdeki çalışmalarda sabit ve değişken uzunlukta çizge büyüklüğü tercih edilebildiği görülmekte, yöntem kararını etkileyen faktörler arasında ise veri seti büyüklüğü, müzikal içerik türü gibi değişkenler bulunmaktadır [50][52]. Çizge büyüklüğü, veri setindeki müzikal içeriklerin müzikal ölçü sayı ortalaması dikkate alınarak belirlenmiştir. Çizge büyüklüğü belirlenirken dikkat edilen iki faktör müzikal kaliteyi korumak ve hesaplama karmaşıklığı sınırlamak olmuştur. Müzikal ölçü üretimi probleminde kullanılan üç değişkenin (kaynak düğüm, hedef düğüm, ilişki tipi) her biri kategorik değişkenler olarak kullanılmıştır.

Şekil 3.5, dört farklı müzikal içerik için müzikal ölçü ilişkilerinin çizge veri yapısı ile gösterimini görsel olarak sunmaktadır. Görsellerde kırmızı renk tekrar ilişkisini, sarı renk transpozisyon ilişkisini, mavi renk melodi ilerlemesi ilişkisini ve yeşil renk ise ritmik dizilim ilişkisini ifade etmektedir. Parçalar arasında gerek müzikal ölçü ilişki sayısı gerekse farklı ilişki türlerine sahip olma durumu değişkenlik gösterebilmektedir. Şekil 3.5'te sol üst kısımda (a)

görülen müzikal içerik üç farklı müzikal ilişkiyi aynı anda içerirken, sağ üst kısımda (b) görülen müzikal içerik başlangıçtaki ve bitişteki ölçülerin tekrar ettiğini göstermektedir. Sol alt kısımda (c) görülen müzikal içerik ise yüksek oranda tekrar ilişkisi içermektedir. Sağ alt kısımda (d) görülen müzikal içeriğin oldukça az ölçüsü arasında ilişki bulunduğu görülmektedir.

Parçalar tekil olarak incelendiğinde, her parçanın baskın bir veya birkaç ilişki içerdiği görülmektedir. Müzikal içeriklerden (a) için baskın ilişki transpozisyon olarak öne çıkarken, (b) ve (c) için tekrar ilişkisinin sıklıkla kullanıldığı, (d) için ise ritmik ilerleme ilişkisinin oldukça yüksek oranda tercih edildiği görülmektedir.

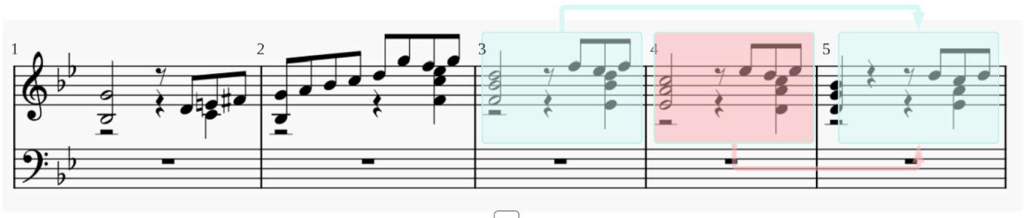


Şekil 3.5: Müzikal ölçü ilişkilerinin çizge veri yapısı ile gösterimi

3.3.1.2. Müzikal Ölçü Koşullu Nota Seviye Gösterim (Sıralı)

Müzikal ölçü koşullu nota seviye gösterim de koşulsuz nota seviye gösterim gibi sıralı gösterim kullanmaktadır. Farkı ise koşulsuz gösterime ek olarak, müzikal ölçü seviyesindeki ilişki bilgisini de gösterime entegre ederek nota seviyesindeki içerik üretimini müzikal ölçü ilişkisi

ile koşullamasıdır. Şekil 3.6'te temsili bir müzikal içerik ve sıralı gösterim matrisi görsel olarak sunulmaktadır. Beş müzikal ölçülük kesitten oluşan örnekte, turkuaz renk ile temsil edilen üçüncü ve beşinci ölçüler arasında transpozisyon ilişkisi olduğu görülmektedir. Beşinci ölçüde bulunan notalar, kendinden önce gelen dördüncü müzikal ölçü notaları ve ilişkide olduğu üçüncü müzikal ölçü notaları ile birlikte temsil edilmektedir. Görselin alt kısmında sunulan matriste ilk dört satır sıralı olarak geçmiş ölçüden gelen bilgileri, son dört satır ise müzikal ölçü ilişkisel olarak elde edilen bilgiyi içermektedir. Geçmiş müzikal ölçüden gelen bilgiler için ilişki sınıfı yeni bir ilişki türü olarak çözüme entegre edilmiştir. Girdi vektörlerinde bulunan nümerik değişkenler, Bölüm 3.2.1'de detaylı açıklaması verilen Min-Max normalizasyon tekniği ile normalize edilmiştir.



3. Ölçü - 5. Ölçü Transpozisyon

Ölçü koşullu sıralı nota seviye gösterim

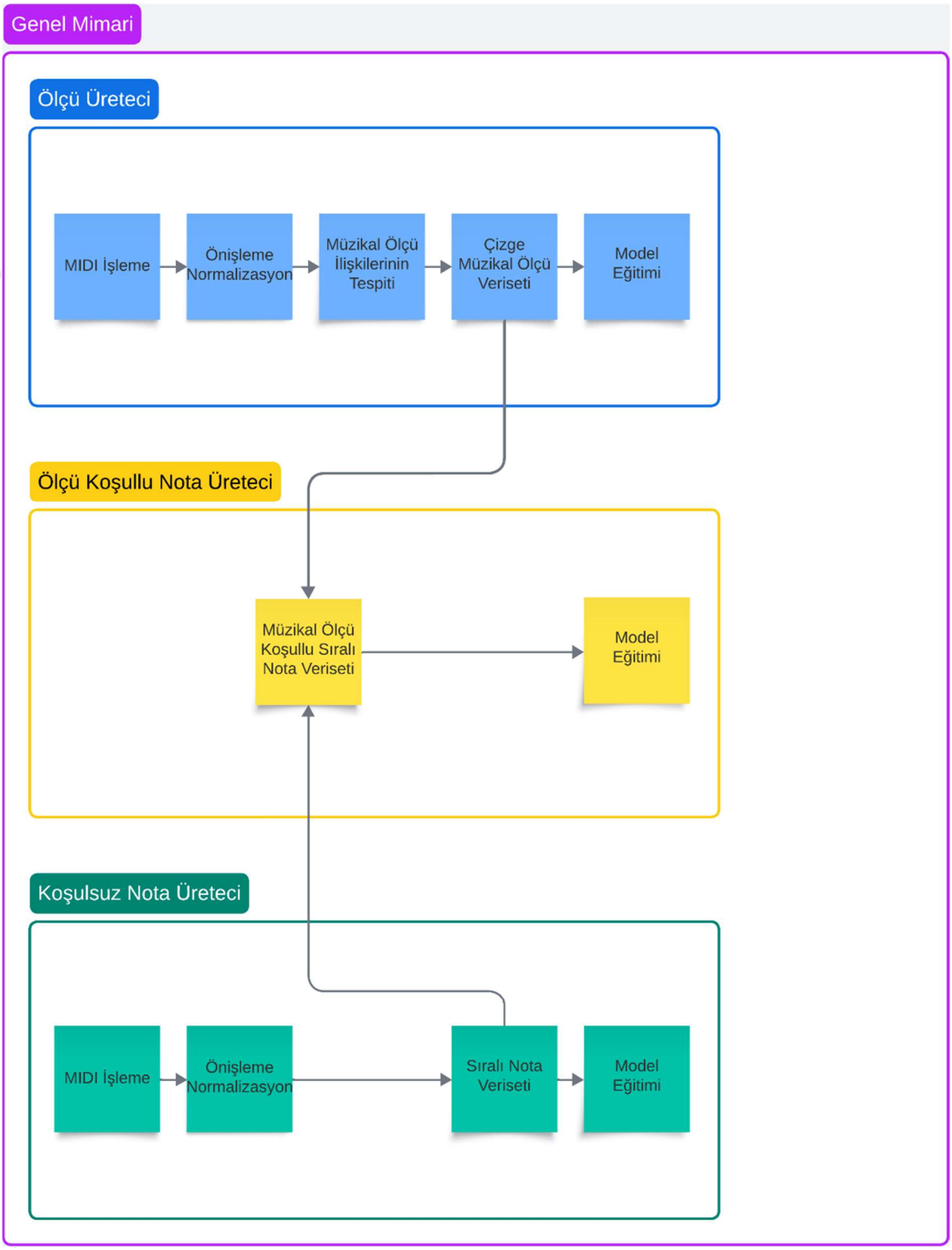
Dizilimsel Nota	int	75	67	70	74	77	43	46
Dizilimsel Çalım Zamanı	float	0.5	0.5	0.5	0.5	0.5	1	1
Dizilimsel Ses Seviyesi	int	70	70	70	70	70	70	70
Dizilimsel İlişki Tipi	int	7	7	7	7	7	7	7
Ölçü Nota	int	75	67	70	74	77	44	48
Ölçü Çalım Zamanı	float	0.5	0.5	0.5	0.5	0.5	1	1
Ölçü Ses Seviyesi	int	75	75	75	75	75	75	75
İlişki Tipi	int	2	2	2	2	2	2	2

Şekil 3.6: Müzikal ölçü koşullu nota seviye gösterim

3.3.2. Mimari Tasarım

GAMA-DL yöntem mimarisi Şekil 3.7'da sunulmaktadır. GAMA-DL ile melodi üretiminde ana unsur nota üretimi problemini yüksek seviyeli müzikal ölçü ilişkileriyle desteklemektir. Mimari, müzikal ölçü üreteci, koşulsuz melodi üreteci ve müzikal ölçü koşullu melodi üreteci olmak üzere üç farklı DÖ yönteminin birleşiminden oluşmaktadır. Müzikal ölçü üreteci yöntemin ilk aşaması olup, melodik yapı içerisindeki yüksek seviyeli ilişkileri üretmektedir. Koşulsuz nota üreteci, nota seviyesinde geçmiş notaları dikkate alarak nota üretiminde görev almaktadır. GAMA-DL mimarisinde koşulsuz nota üretici olarak MTHA-LSTM yöntemi kullanılmıştır. Koşullu nota üretici yöntem ise ilişkili müzikal ölçü ile geçmiş müzikal ölçüleri aynı anda değerlendirerek yeni müzikal ölçüyü nota seviyesinde oluşturan DÖ yöntemini ifade

etmektedir. Üç aşamada da referans modeller RNN ve LSTM, önerilen yöntem ise MTHA-LSTM olarak kullanılmıştır.

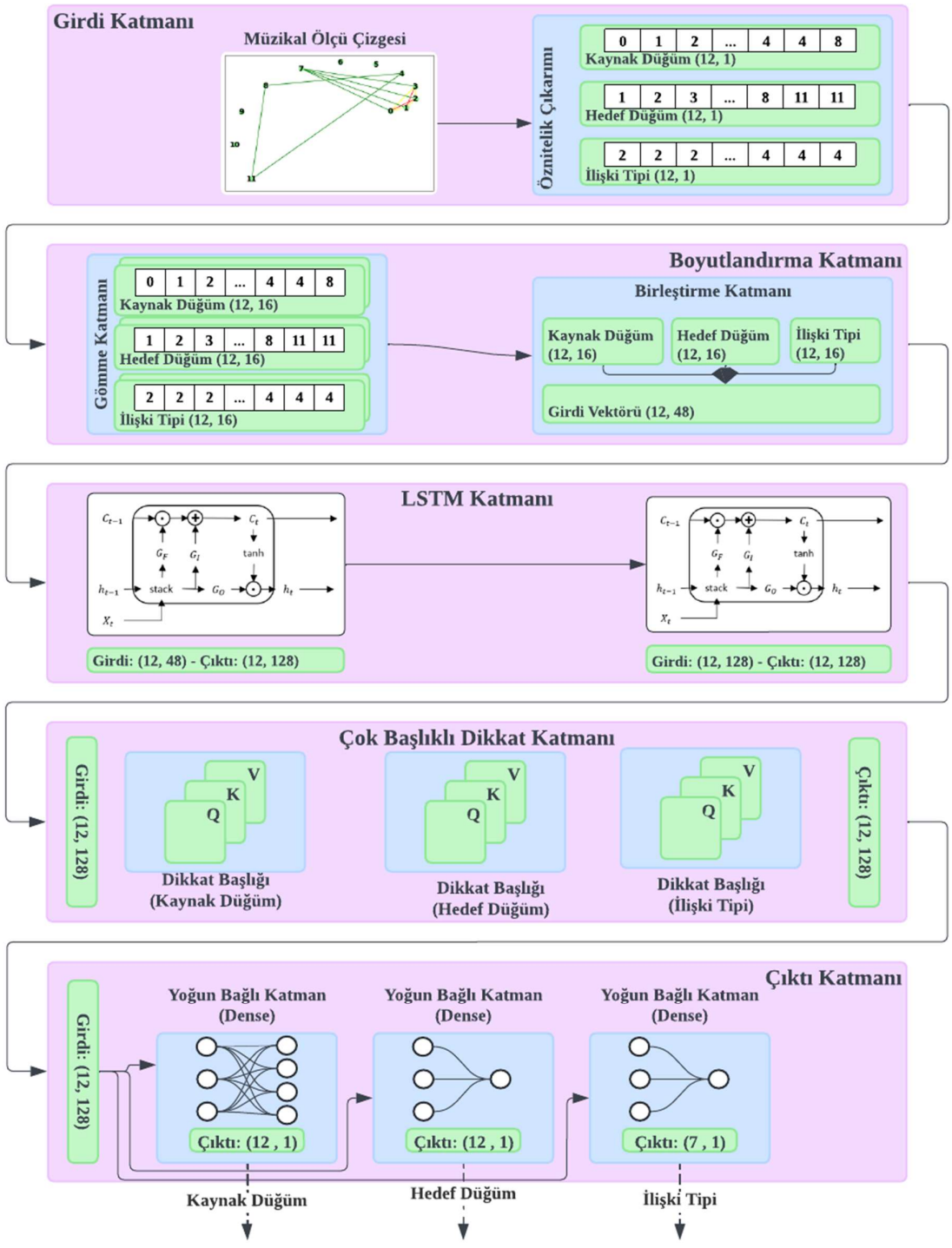


Şekil 3.7: Melodi üretici modül mimari yapısı

3.3.2.1. Ölçü Üretimi

GAMA-DL yönteminin müzikal ölçü seviyesindeki ölçü ilişkilerini yakalama ve yapay olarak üretme görevini, ölçü üretici model üstlenmektedir. Tez çalışmasında müzikal ölçülerin temsilinde çizge gösterimi kullanılmıştır. Müzikal ölçü üretimi problemi, çizge üretim problemi bağlamında incelenmektedir. Çizge üretimi problemi, düğüm sayısı önceden belirlenmiş bir çizgedeki düğüm komşuluklarının belirlenmesi problemi olarak çözümlenmektedir. Çizgelerde komşuluklar, komşuluk listesi gösterimi ile temsil edilmiştir. Ölçü komşulukları, "kaynak düğüm", "hedef düğüm" ve "ilişki tipi" olmak üzere üç değişken ile temsil edilmektedir. Kaynak düğüm ve hedef düğüm değişkenleri düğümlerin çizgedeki konumunu ifade ederken, ilişki tipi ise kategorik bir değişken ile ilişki tiplerini ifade etmektedir.

Şekil 3.8, önerilen ölçü üretici DÖ yönteminin mimari tasarımını sunmaktadır. Ölçü üretici yöntem, Bölüm 3.2.2'de detaylı açıklaması verilen MTHA-LSTM yönteminin çizgelerde komşuluk üretimi problemine uyarlanmış halidir. Girdi olarak müzikal çizge verisetleri kullanılmakta ve üç değişken için ayrı ayrı girdi vektörleri elde edilmektedir. Çıktı olarak ise her bir adımda çizgedeki bir komşuluk ilişkisi belirlenmektedir.



Şekil 3.8: Müzikal ölçü üretici modül mimari tasarımı

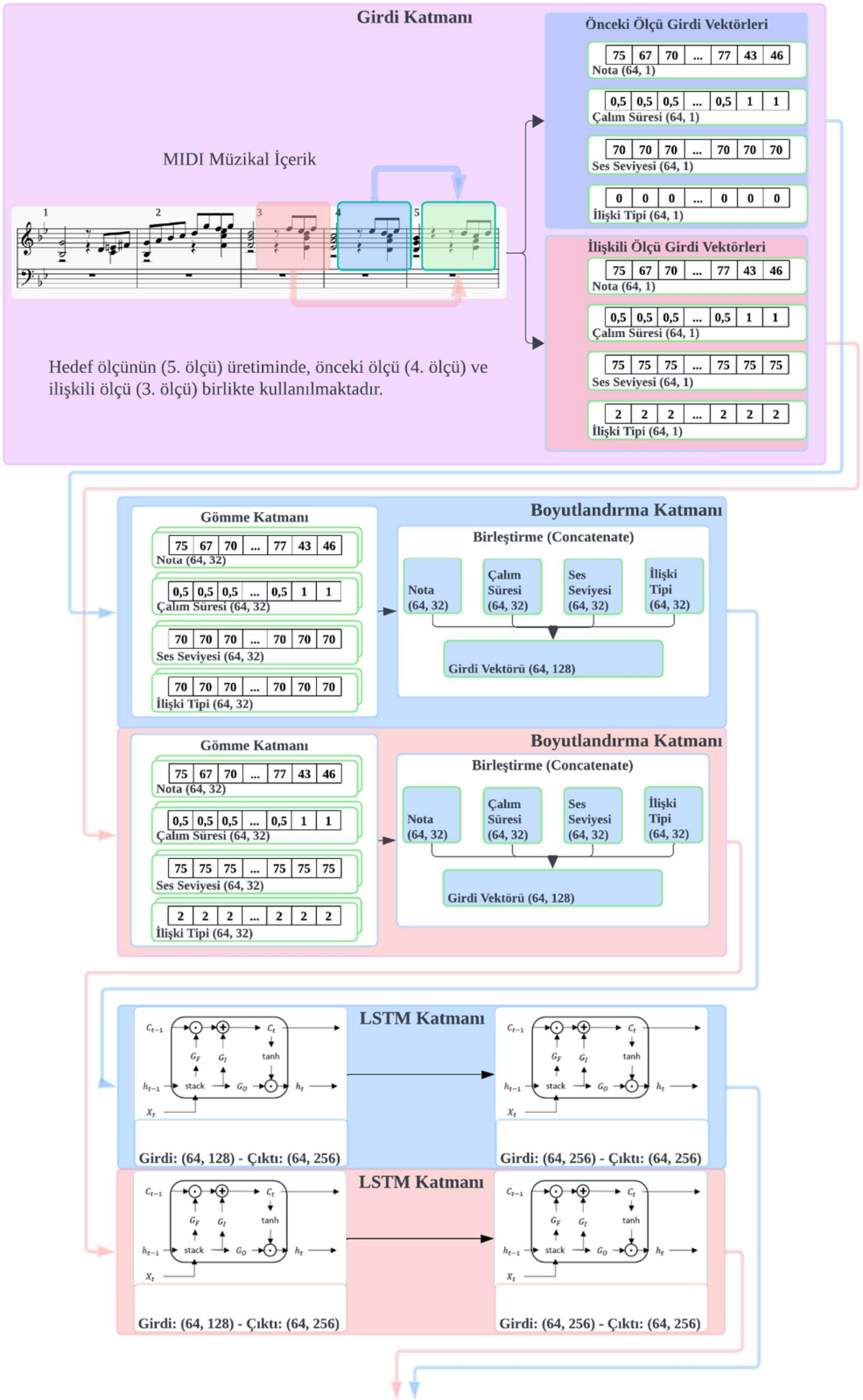
3.3.2.2. Müzikal Ölçü Koşullu Nota Üretimi

Müzikal ölçü koşullu nota üretici modül, önceki müzikal ölçü ve ilişkili müzikal ölçü notalarını birlikte kullanarak, sonraki müzikal ölçü notalarını üretmek amacıyla kurgulanmıştır.

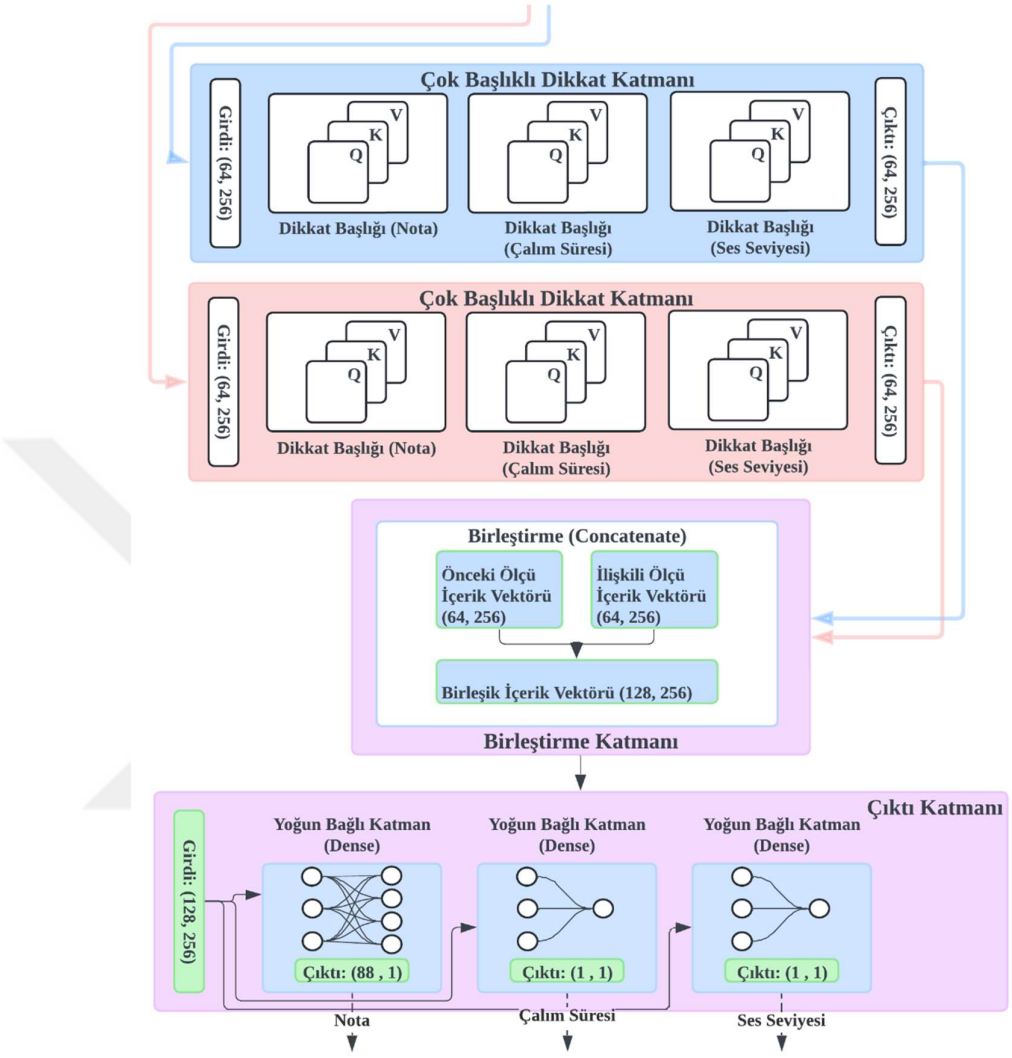
Geliştirilen modelin mimari tasarımı Şekil 3.9 ve Şekil 3.10'da sunulmaktadır. Görsel sunumda mavi tonda ifade edilen bloklar önceki müzikal ölçü vektörlerinin işlendiği katmanları, kırmızı tonda ifade edilen bloklar ilişkili müzikal ölçü vektörlerinin işlendiği katmanları, mor tonda ifade edilen bloklar ise ortak katmanları ifade etmektedir.

Yöntem bir önceki ölçü ve bir ilişkili ölçü olmak üzere iki ölçünün nota seviye girdi vektörlerini kabul etmektedir. Her iki ölçü için de "nota", "çalım süresi", "ses seviyesi" ve "ilişki tipi" olmak üzere dört farklı girdi vektörü, girdi katmanında elde edilmektedir. Önceki ölçüden ve ilişkili ölçüden gelen bu girdi vektörlerinin her ikisi de MTHA-LSTM başlığında önerilen yöntemle fakat birbirinden soyut iki ayrı kanalda işlenmektedir. Her iki ölçünün notaları için iki ayrı içerik vektörü (64, 256) ebatlarında elde edilmektedir. Bu içerik vektörleri ise en son bir birleştirme aşamasında birleştirilmekte ve (128, 256) ebatında bir vektör olarak çıkış değerlerini üretmek üzere yoğun bağlantılı katmana iletilmektedir.

Yöntemin çıkışı ise yeni ölçünün nota, çalım süresi ve ses seviyesi vektörleri için toplamda üç vektör üretmektedir. Değişkenlerin farklı karakteristik özellikleri bulunduğundan, her çıkış değişkeni için ayrı bir yoğun bağlı katman kullanılmaktadır.



Şekil 3.9: Müzikal ölçü koşullu nota üretici modül mimari tasarımı



Şekil 3.10: Müzikal ölçü koşullu nota üretici modül mimari tasarımı (Devamı)

3.4. Eşlik Üretimi

Tek-sesli melodileri çok-sesli müzikal içeriklere dönüştürebilmek için müzisyenlerin kullandığı akor eşliği (chord accompaniment), bas eşliği (bass accompaniment) veya karşıt melodi (counterpoint) gibi teknikler bulunmaktadır [52]. Bu yaklaşımların her biri, ayrıca incelenebilecek ve farklı müzikal nüanslar oluşturma kapasitesine sahip etkili yöntemlerdir. Bu çalışmada, en temel yaklaşımlardan olup, popülerliği ve uygulama kolaylığı sağlaması açısından akor eşliği yaklaşımı kullanılmıştır. Akor eşliği yaklaşımında, melodiyi destekleyen çok-sesli akor dizileri elde edilmesi amaçlanmaktadır.

Müzikal eşlik üretimi, bir melodinin armonik yapısına uygun olarak akorlar elde edilmesi amacıyla gerçekleştirilen sürece verilen isimdir. Bu süreç, müzikal teori kurallarına dayanarak algoritmik olarak gerçekleştirilebilmektedir. Melodinin hangi anahtarda (tonalite) olduğunu belirlemek, eşlik üretimi için uygun akorların seçiminde ilk adımı oluşturmaktadır. Melodinin başlangıç ve bitiş notaları, tonalitenin belirlenmesinde kullanılmıştır. Bir tonalitede kullanılabilecek akorların belirli bir diziyi takip etmesi beklenmektedir. Örneğin, C majör tonalitesinde, kullanılabilecek ana akorlar şöyledir:

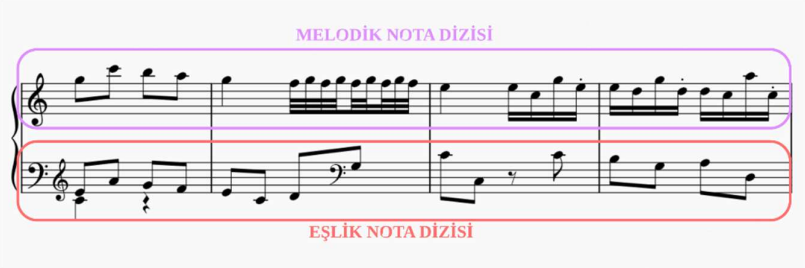
- C-majör (C-E-G)
- D-minör (D-F-A)
- E-minör (E-G-B)
- F-majör (F-A-C)
- G-majör (G-B-D)
- A-minör (A-C-E)
- B-dim (B-D-F)

Eşlik üretim algoritması, melodi m'yi girdi olarak almakta ve eşlik z'yi çıktı olarak sunmaktadır (Algoritma 3.1). Melodi girdisindeki notaların (M_t), n_c sözlüğü aracılığıyla uygun akorları (b) belirlenmektedir. Sözlükte her notaya karşılık akor sınıfları bulunmaktadır. Ardından notaya karşılık gelebilecek uygun akorlar arasından bir tanesi rassal fonksiyonu aracılığıyla seçilmektedir. Örneğin, melodide bulunan bir C notası için, C-majör, A-minör ve F-majör akorlarından biri eşlik akoru olarak atanabilmektedir. Ardından seçilen akor, eşlik dizisine eklenmekte ve tüm akorların seçimi tamamlandığında, eşlik üretimi tamamlanmaktadır.

Algoritma 3.1: Eşlik Üretimi

Gerekenler: M, Melodik nota dizisi.
$n_c \leftarrow \{C, D, E, F, G, A, B\}$
Z ← eşlik dizisi
b ← baz melodik nota
c ← seçilen eşlik akor
for t = note in M **do**
  b ← $n_c(M_t)$
  $c \leftarrow rassal()$
  $z_t \leftarrow$ c
end for
return Z

Şekil 3.11, melodik nota dizisinin eşlik nota dizisi ile zenginleştirildiği temsili bir müzikal bir içerik sunmaktadır. Melodik dizi tek-sesli bir içerik ihtiva ederken, eşlik dizisi çok-sesli bir yapıya sahiptir.



Şekil 3.11: Melodi ve eşlik nota dizileri ile sembolik gösterimde temsili müzikal içerik

4. BULGULAR

Bu bölümde MTHA-LSTM ve GAMA-DL yöntemlerinin nicel ve nitel ölçümleri ve önerilen yöntemlerin literatürdeki çalışmalar ve baz yöntemler ile karşılaştırmalı incelenmeleri sunulmaktadır. Nicel ölçümler, minimum karesel hata (MSE), nota, çalım süresi ve ses seviyesi gibi değişkenler için Özgün Değer Sayısı (ÖDS), ortalama, varyans, tekrarlı ifade sayısı (TİS) ve kosinüs benzerliği (KB) metrikleri ile değerlendirilmektedir. Nitel ölçümler, MTHA-LSTM ve GAMA-DL yöntemlerinin anket ölçüm yöntemiyle değerlendirildiği analizleri içermektedir.

4.1. Nicel Ölçüm Yöntemleri

Nicel deneylerde amaç, üretici modellerin ortaya çıkardığı içeriklerin gerçek müzikal içeriklerin istatistiki özelliklerine yakınsayabilme durumlarını ölçmektir. Üretici modeller ile üretilen yapay müzikal içeriklerin nicel yöntemler ile ölçümü, literatürde çözümü üzerinde fikir birliğine varılamamış ve çözülmeyi bekleyen önemli araştırma problemlerinden biridir. Yaygın kullanımda model verimliliğinin belirlenmesinde doğal dil işleme alanındaki nicel yöntemlerin sıklıkla kullanıldığı görülmekte olup, müzikal metriklerin geliştirilmesine ve bu alanda kullanım birliği sağlanmasına yönelik çalışmalara ihtiyaç bulunmaktadır. Çalışmada nota ve müzikal ölçü seviyesinde farklı metrikler ile ölçümler gerçekleştirilmiştir.

Model eğitimi aşamasında performans değerlendirmesi amacıyla regresyon modellerinde yaygın olarak kullanılan minimum karesel hata (MSE) metriği tercih edilmiştir (4.1). MSE, gerçek değer ile DÖ yöntemi tarafından tahminlenen değerlerin arasındaki farkların karelerinin ortalamasıdır. Elde edilen değer her zaman pozitif ya da sıfır olarak elde edilmektedir, bu değer ne kadar düşükse model performansı o kadar iyi gerçekleşmiştir. MSE, hata ölçümlerinin karesi üzerinden hesaplandığından, büyük hatalara daha fazla ağırlık verme eğilimindedir. Çalışmada MSE metriğinin tercih edilmesinin sebebi, aykırı değerlere duyarlı olmasıdır.

MSE metriği hesaplanırken n adet tahmin değer $\acute{y}_i$ ve gerçek değer y_i verildiği varsayılsın. Öncelikle her bir tahmin için, gerçek değer ile arasındaki hata hesaplanır. Daha sonra bu hatanın karesi alınır. Son olarak tüm hata kareleri toplanarak toplam değer sayısına bölünür.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\acute{y}_i - y_i)^2 \tag{4.1}$$

Nota seviyesindeki öznitelikler olan nota, çalım süresi ve ses seviyesi için ÖDS ve varyans metrikleri kullanılmıştır. ÖDS metriği, bir modelin ürettiği müzik parçasındaki özgün değerlerin sayısının ölçülmesini ifade etmektedir. ÖDS modelin yaratıcılığını ve orijinalliğini değerlendirmek için kullanılmaktadır. YZ modelleri için amaç, gerçek müzikal içeriklerin ÖDS değerlerine yakınsamaktır. Varyans ise notaların dağılımı hakkında bilgi sunmakta ve müzikalite hakkında genel bir bilgi sunmaktadır. Varyans ve ÖDS için düşük değerler elde edilmesi, modelin özgün nota üretmede zorluk yaşadığına, yüksek değerler elde edilmesi ise ritmik ve melodik yapının kurgulanmasında zorlanıldığına işaret etmektedir. Buna göre varyans ve ÖDS metriklerinde gerçek içeriklere yakınsamak, model için başarı ifade etmektedir.

Ölçü seviyesindeki başarı metrikleri, literatürde sıkça kullanılan TİS ve KB metrikleridir [53]. TİS metriği, bir müzik parçasındaki tekrar eden motiflerin veya kalıpların sayısını ölçmektedir. Bu motifler, melodik figürler, armonik ilerlemeler veya ritmik desenler gibi çeşitli unsurları içerebilmektedir. Tekrar eden motiflerin sayısı, bir pencere boyutu (L) ve eşik değeri (m) kullanılarak hesaplanmaktadır (4.2). Pencere boyutu, bir motifin tekrar olarak kabul edilmesi için en az kaç kez arka arkaya çalınması gerektiğini göstermektedir. Eşik değeri (m), bir motifin benzer olarak kabul edilmesi için ne kadar benzer olması gerektiğini belirlemektedir.

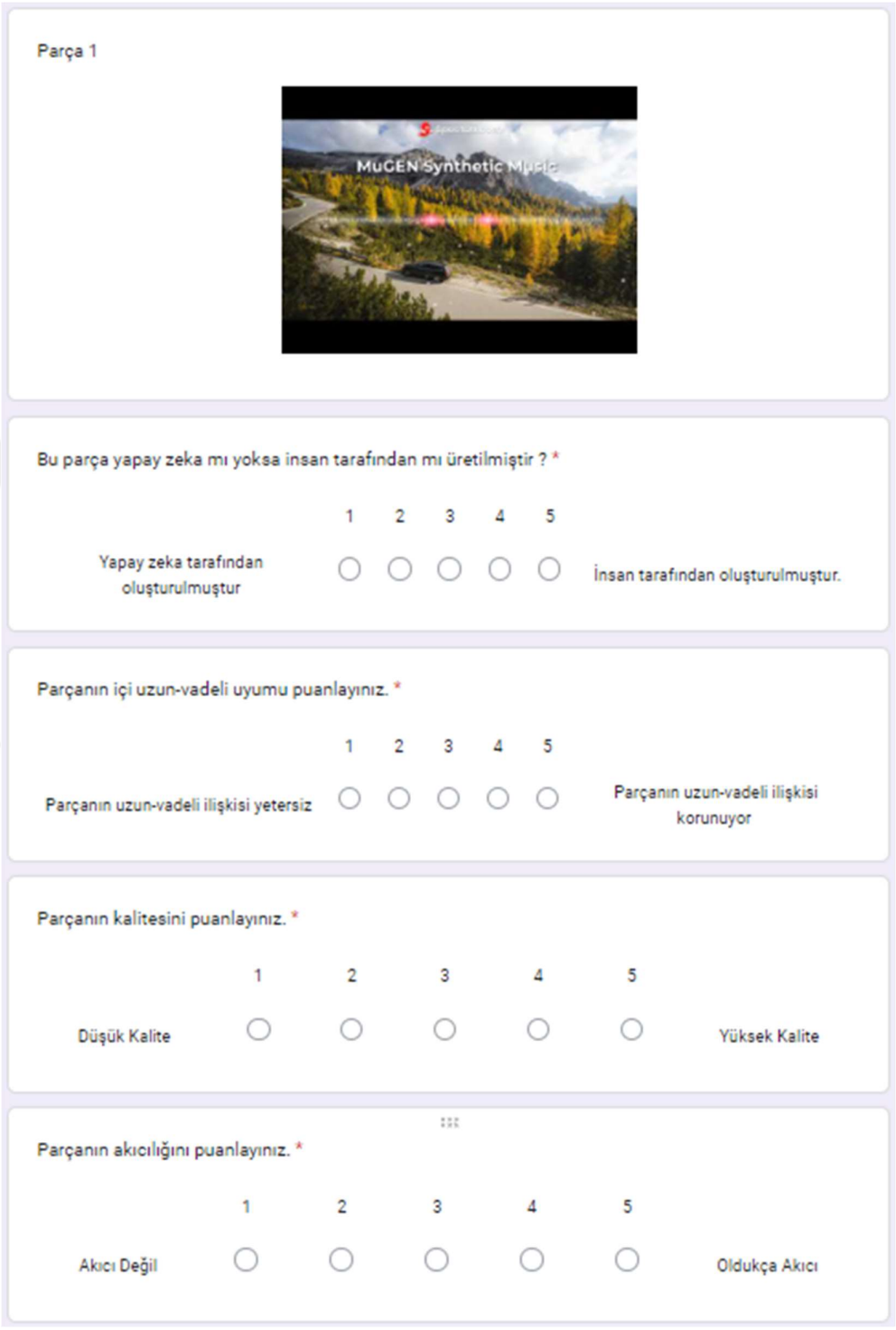
$$\mathrm{TİS}\ (M, L, m) = \sum \frac{(\mathrm{m} \in \mathrm{M})\ |\mathrm{m}|}{L} \tag{4.2}$$

KB matrisi, iki vektör arasındaki açısal benzerliği ölçen bir metriği temsil etmekte ve metin madenciliği, doğal dil işleme ve müzik analizi gibi alanlarda kullanılmaktadır. Bu metrik, iki vektörün iç çarpımının, bu vektörlerin normlarının çarpımına bölünmesi ile elde edilmektedir. Matematiksel olarak, iki A ve B müzikal ölçü vektörü arasındaki kosinüs benzerliği (4.3)'deki gibi hesaplanmaktadır.

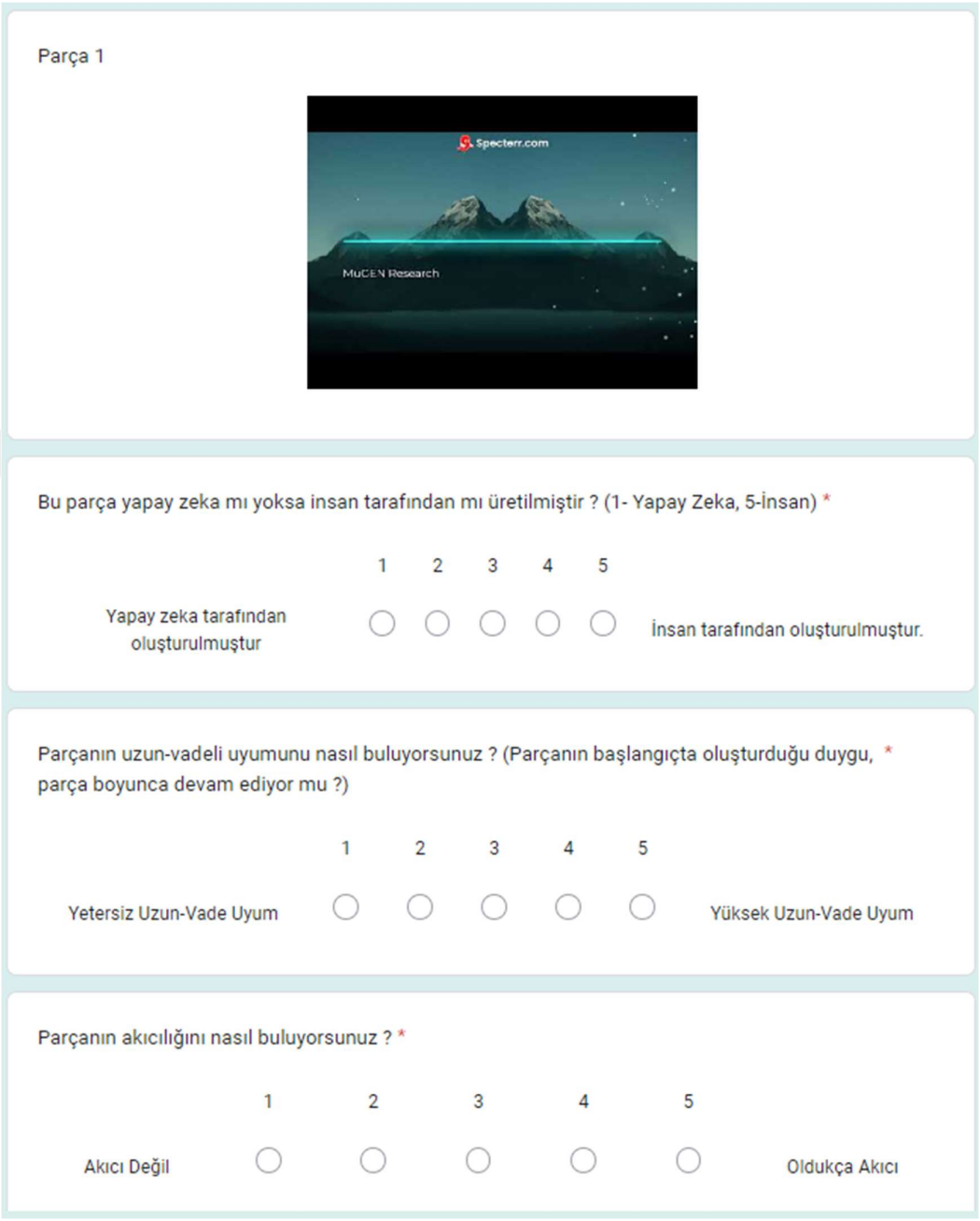
$$\cos(\theta) = \frac{A \cdot B}{||A||\,||B||} \tag{4.3}$$

4.2. Nitel Ölçüm Yöntemleri

Müzikal içeriklerin sanatsal yapıları gereği nicel ölçüm yöntemlerinin tek başına kullanımı yanıltıcı olabilmektedir. Literatürdeki çalışmalar, nicel ölçümlerin nitel ölçümlerce desteklendiği hibrit bir değerlendirmenin gerekliliğini ortaya koymaktadır [54]. Tez kapsamında geliştirilen MTHA-LSTM ve GAMA-DL yöntemlerinin her ikisinde de nitel ölçüm yöntemi olarak anket araştırması gerçekleştirilmiştir. Her iki anket araştırması da en az 50 katılımcı ile gerçekleştirilmiştir. Araştırmalarda, Google Forms platformu kullanılmıştır. Anketlerde kullanıcılara gerçek ve YZ tarafından üretilmiş müzikal içerikler karışık sırada dinletilmiştir. MTHA-LSTM kullanıcı araştırmasında soru örneği Şekil 4.1'de, GAMA-DL kullanıcı araştırmasında soru örneği ise Şekil 4.2'de görülmektedir.



Şekil 4.1: MTHA-LSTM kullanıcı araştırması soru örneği



Şekil 4.2: GAMA-DL kullanıcı araştırması soru örneği

Soruların skorlanmasında Likert ölçeği kullanılmıştır [55]. Likert ölçeği, anket araştırmalarında yaygın olarak kullanılan bir psikometrik ölçümdür. Ölçek, katılımcıların belirli ifadelere veya sorulara katılma veya katılmama düzeylerini belirtmeleri için kullanılmaktadır. Literatürde genellikle beş veya yedi puanlı bir derecelendirme ölçeğinin tercih edildiği görülmektedir [55]. Tez kapsamında düzenlenen kullanıcı araştırmalarında beş puanlı Likert ölçeği tercih edilmiştir. Beş puanlı bir Likert ölçeğinin bir örneği şöyledir:

- 1: Kesinlikle Katılmıyorum
- 2: Katılmıyorum
- 3: Kararsız
- 4: Katılıyorum
- 5: Kesinlikle Katılıyorum

MTHA-LSTM çalışmasında kullanıcı araştırması, bir gerçek parça ile üç YZ üretimi parçanın dinleme testleri ile gerçekleştirilmiştir. Her parça için sorulan sorular şu şekildedir:

- Bu parça bir insan mı yoksa yapay zekâ tarafından mı bestelenmiştir?
- Parçanın melodik yapısını 1-5 arasında değerlendiriniz.
- Parçanın kalitesini 1-5 arasında değerlendiriniz.
- Parçanın akıcılığını 1-5 arasında değerlendiriniz.

GAMA-DL çalışmasında ise kullanıcı araştırması, iki gerçek, iki RNN üretimi, iki LSTM üretimi, iki MTHA-LSTM üretimi ve iki GAMA-DL üretimi olmak üzere toplam on parça ile gerçekleştirilmiştir. Her parça için katılımcılara üç soru yöneltilmiştir:

- Bu parça yapay zekâ mı yoksa insan tarafından mı üretilmiştir? (1- Yapay Zekâ, 5- İnsan)
- Parçanın uzun-vadeli uyumunu nasıl buluyorsunuz? (Parçanın başlangıçta oluşturduğu duygu, parça boyunca devam ediyor mu?)
- Parçanın akıcılığını nasıl buluyorsunuz?

4.3. MTHA-LSTM Sonuçları

Bu kısımda tez araştırması kapsamında geliştirilen ilk yöntem olan MTHA-LSTM yönteminin nicel ve nitel ölçüm sonuçları sunulmaktadır. Deneylerde gerçek müzikal içeriklerin sonuçları referans kabul edilerek, baz yöntemler olan “RNN”, “LSTM” ile önerilen “MTHA-LSTM” yönteminin müzikal içerik üretimindeki etkinlikleri karşılaştırmalı olarak incelenmiştir.

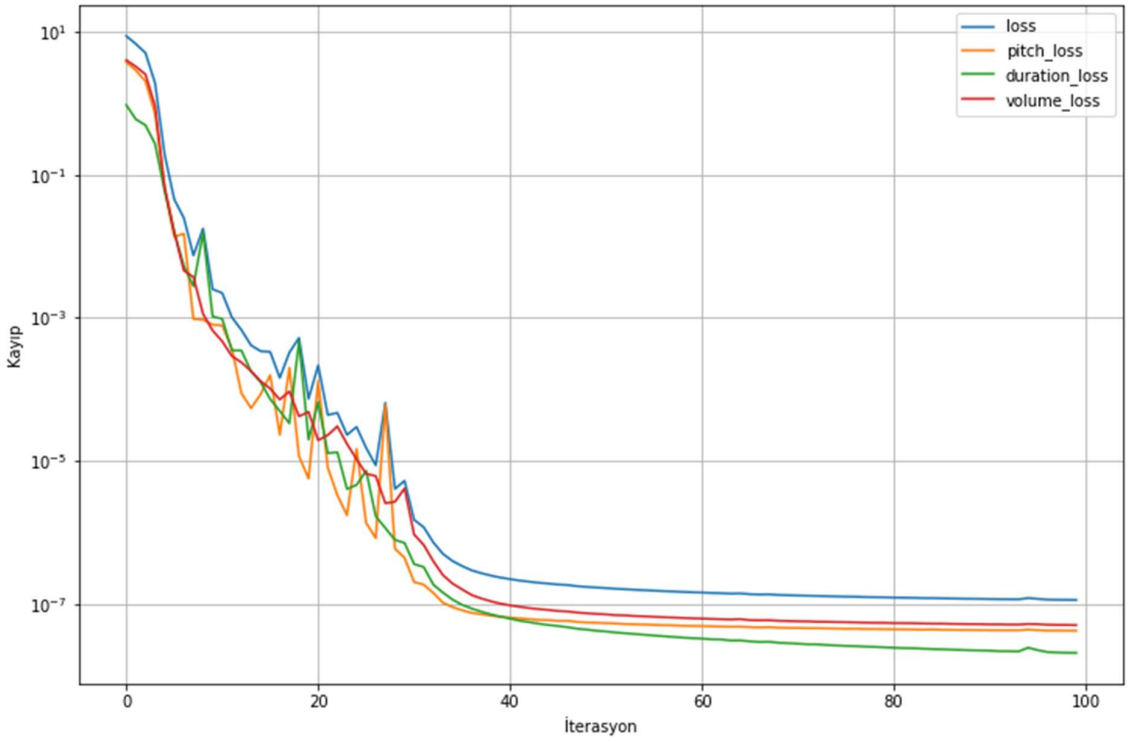
4.3.1. Nicel Ölçümler

Tablo 4.1, MTHA-LSTM ile baz modellerin eğitim kayıp performanslarına ilişkin nicel deneysel sonuçlarını sunmaktadır. Dikkat ağlarının kullanımının çözüm uzaylarında eğitim gelişimini önemli ölçüde iyileştirmekte ve MTHA-LSTM'in gerçek bilgi uzayına hızlı yakınsama sağlamaktadır. RNN ve LSTM yöntemlerinin ise özellikle nota değişkeni için öğrenme başarımının oldukça düşük kaldığı görülmektedir.

Tablo 4.1: YZ model öğrenme başarımları

Yöntem	MSE (Toplam)	MSE (Çalım Süresi)	MSE (Ses Seviyesi)	MSE (Nota)
RNN	1.8430	0.0245	0.0702	2.9936
LSTM	1.8822	0.0126	0.0684	3.0794
MTHA-LSTM	**9,58E-03**	**4,26E-04**	**8,52E-03**	**6,40E-04**

Şekil 4.3'de MTHA-LSTM yöntemi için genel kayıp eğrisi ile öznitelik bazında kayıp eğrileri birlikte sunulmaktadır. Kayıp eğrileri, ilk iterasyonlarda görülen keskin düşüşün, devam eden iterasyonlar boyunca düzenli bir iyileşme ile desteklendiğini göstermektedir. Değişkenlerin tekil kayıp eğrilerine bakıldığında, kategorik nota değişkeninin en yüksek hesaplama karmaşıklığını oluşturduğu görülmektedir. Çalım süresi ve ses seviyesini temsil eden diğer öznitelikler sayısal değişkenlerden oluştuğundan, nota özniteliğine göre MTHA-LSTM yönteminin toplam nöron kapasitesinin çok daha az bir kısmı ile modellenebilmektedir.



Şekil 4.3: Model Eğitim Aşaması Kayıp Eğrileri

Müzikal içeriklerdeki önemli değişkenlerden biri de ritmik ve melodik yapının kavranabilmesidir. NLP'de sıklıkla kullanılan TİS metriği, müzikal içeriklerde de melodik ve ritmik yapının ölçümünde sıklıkla başvurulan bir metrik olarak öne çıkmaktadır. Çalışmada, TİS metriği modellerin eğitimi sırasında kullanılmamış olup, eğitim sonrası ölçüm metriği olarak değerlendirilmiştir. Tablo 4.2'de gerçek müzikal içerikler referans alınarak, MTHA-LSTM yönteminin başarımı baz yöntemler RNN ve LSTM ile karşılaştırılmıştır. Varyans ve ÖDS metrikleri nota seviyesinde özgün ve uygun istatistiki dağılımda içerik üretimini ölçerken TİS metriği ile ritmik yapıya uygunluk kontrol edilmektedir. TİS analizi, sadece nota değişkeni üzerinde gerçekleştirilmiştir. MTHA-LSTM yöntemi tüm metriklerde insan bestesi içeriklere en yakın sonuçları sunarak baz yöntemlerin üzerinde başarım sunmuştur. Dikkat tabanlı yöntemlerin uzak değişkenler arasındaki ilişkileri temsil etmekte verimi oldukça artırdığı görülmektedir.

Tablo 4.2: MTHA-LSTM nicel ölçüm sonuçları

Model	Varyans			ÖDS			TİS	
	Nota	**Çalım Süresi**	**Ses Şiddeti**	**Nota**	**Çalım Süresi**	**Ses Şiddeti**	**2 ifadeli tekrar**	**3 ifadeli tekrar**
İnsan Bestesi	*27.83*	*0.08*	*32.01*	*21*	*2*	*26*	*21*	*17*
MTHA-LSTM	**30.84**	**0.06**	**28.49**	**30**	**11**	**23**	**24**	**21**
LSTM	67.30	0.03	32.96	29	3	39	27	24
RNN	91.41	0.14	23.84	38	8	56	28	26

4.3.2. Nitel Ölçümler

MTHA-LSTM yöntemi için nitel ölçüm yöntemi olarak kullanıcı araştırması uygulanmıştır. Araştırmada katılımcılar sıradan dinleyicilerden ve profesyonel müzisyen veya teorik altyapıya sahip kişilere kadar iki farklı bilgi düzeyine sahip çeşitli altyapılardan insanlar arasından seçilmiştir. Anket bir hafta boyunca toplamda 57 katılımcının değerlendirmesi ile gerçekleştirilmiştir. Katılımcıların %61,4'ü kendilerini sıradan dinleyici olarak ifade ederken, %38,6'sı teorik veya mesleki müzikal geçmişi olan kişiler olduğunu ifade etmiştir. Tablo 4.3'te sonuçları sunulmakta olan kullanıcı araştırması, üç farklı bakış açısıyla değerlendirilmiştir. Genel puanlar açısından MTHA-LSTM modelinin, üç metrikte gerçeğe en yakın skorları yakaladığı, müzikal kalitede ise oldukça rekabetçi bir skor ortaya çıkardığı görülmektedir. Müzikal açıdan deneyimli katılımcılar için MTHA-LSTM tüm ölçümlerde en yüksek genel puanları elde etmiştir. Sıradan dinleyiciler ise LSTM yönteminin daha akıcı ve insansı olduğunu ifade etmiştir. Bununla birlikte uzun vadeli melodik yapıda MTHA-LSTM yönteminin deneyimden seviyesinden bağımsız olarak da tüm katılımcılar arasında hem gerçeğe daha yakın hem de daha yüksek skorlar aldığı görülmektedir.

Tablo 4.3: MTHA-LSTM Kullanıcı Araştırması Sonuçları

Katılımcı Türü	Yöntemler	İnsan bestesi mi (1-5)	Melodik Yapı (1-5)	Kalite (1-5)	Akıcılık (1-5)
Tüm Dinleyiciler	*İnsan Bestesi*	*3,105*	*3,456*	*3,386*	*3,49*
	MTHA-LSTM	**2,526**	**3,157**	2,614	**2,754**
	LSTM	2,47	2,65	**2,7**	2,65
	RNN	2,39	2,37	2,33	2,16
Sıradan Dinleyici	*İnsan Bestesi*	*3,363*	*3,5*	*3,454*	*3,545*
	MTHA-LSTM	2,454	**3,045**	2,409	2,636
	LSTM	**2,64**	2,59	**2,68**	**2,82**
	RNN	2,59	2,14	1,95	1,73
Eğitimli Müzisyenler	*İnsan Bestesi*	*2,942*	*3,428*	*3,342*	*3,457*
	MTHA-LSTM	**2,571**	**3,228**	**2,742**	**2,828**
	LSTM	2,37	2,69	2,71	2,54
	RNN	2,26	2,51	2,57	2,43

4.4. GAMA-DL Sonuçları

Bu kısımda tez araştırması kapsamında geliştirilen ikinci yöntem olan GAMA-DL yönteminin nicel ve nitel ölçüm sonuçları sunulmaktadır. Deneylerde gerçek müzikal içeriklerin sonuçları referans kabul edilmiştir. GAMA-DL yönteminin müzikal içerik üretimindeki performansı Bölüm 3.2'de önerilen MTHA-LSTM ve baz modeller RNN ve LSTM ile karşılaştırmalı olarak incelenmiştir.

4.4.1. Nicel Ölçümler

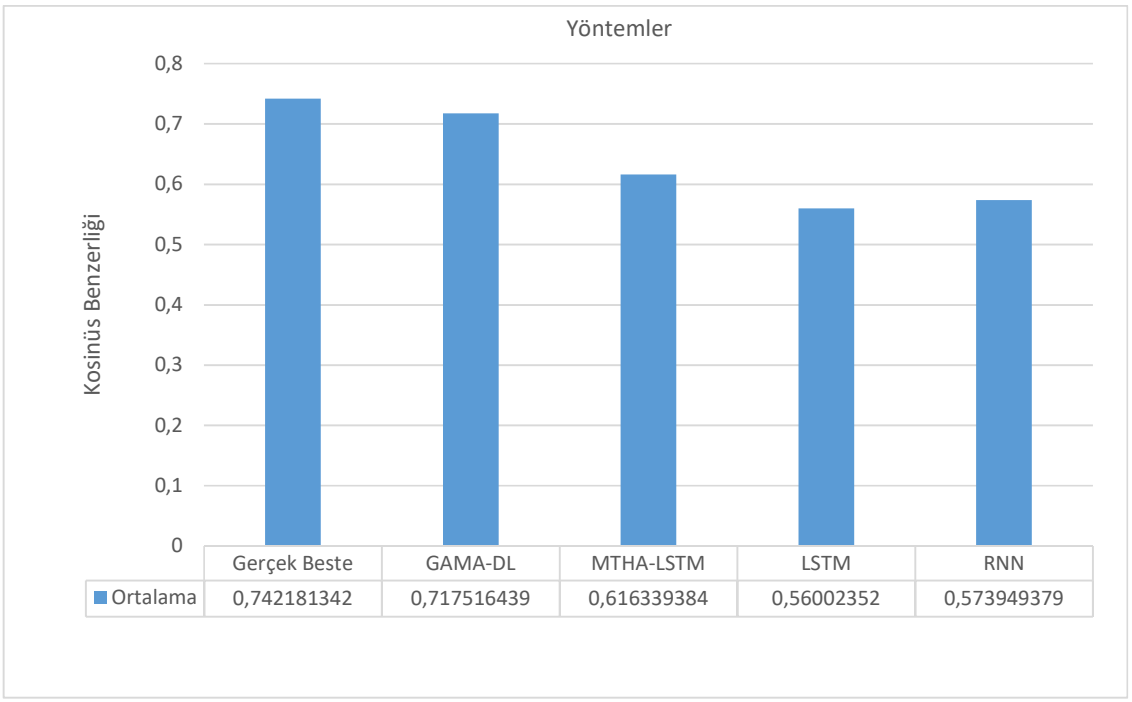
Tablo 4.4, önerilen yöntemler GAMA-DL ve MTHA-LSTM ile referans yöntemler LSTM ve RNN ile üretilen müzikal içeriklerdeki ÖDS ve varyans değerlerini karşılaştırmalı olarak sunmaktadır. İnsan bestesi müzikal içerikler, referans noktası olarak kabul edilmiş ve insan bestesi müzikal içeriklere yakınlık başarım ölçütü olarak kullanılmıştır. Ölçümler nota seviyesinde varyans ve ÖDS metrikleri, müzikal ölçü seviyesinde ise KB metriği ile sınanmıştır. Nota seviyesinde GAMA-DL yönteminin insan bestesi müzikal içeriklere en yakın sonuçları ortaya koyduğu, MTHA-LSTM yönteminin ise bu sonuçlara yakınsayabildiği görülmektedir. Baz yöntemler arasından RNN yüksek varyans göstermekte ve çok sayıda özgün değeri melodik içeriklere eklemektedir. Bu durum bir bağlam ve melodik akış yakalanamadığını göstermektedir. LSTM ise genel olarak RNN gibi yüksek varyans göstermekle birlikte, çalım süresi değişkeninde de çok düşük varyans göstermiştir. Düşük

varyansın sonucu olarak notaların çalım zamanları monoton kalmış, ritmik yapı kurgulanamamıştır.

Tablo 4.4: Varyans ve ÖDS metrikleri ile nicel sonuçlar

Model	Varyans			Özgün Değer Sayısı		
	Nota	Çalım Süresi	Ses Şiddeti	Nota	Çalım Süresi	Ses Şiddeti
İnsan Bestesi	*27.83*	*0.08*	*32.01*	*21*	*2*	*26*
GAMA-DL	**26.96**	**0.07**	**31.76**	**22**	**4**	**25**
MTHA-LSTM	30.84	0.06	28.49	30	11	23
LSTM	67.30	0.03	32.96	29	3	39
RNN	91.41	0.14	23.84	38	8	56

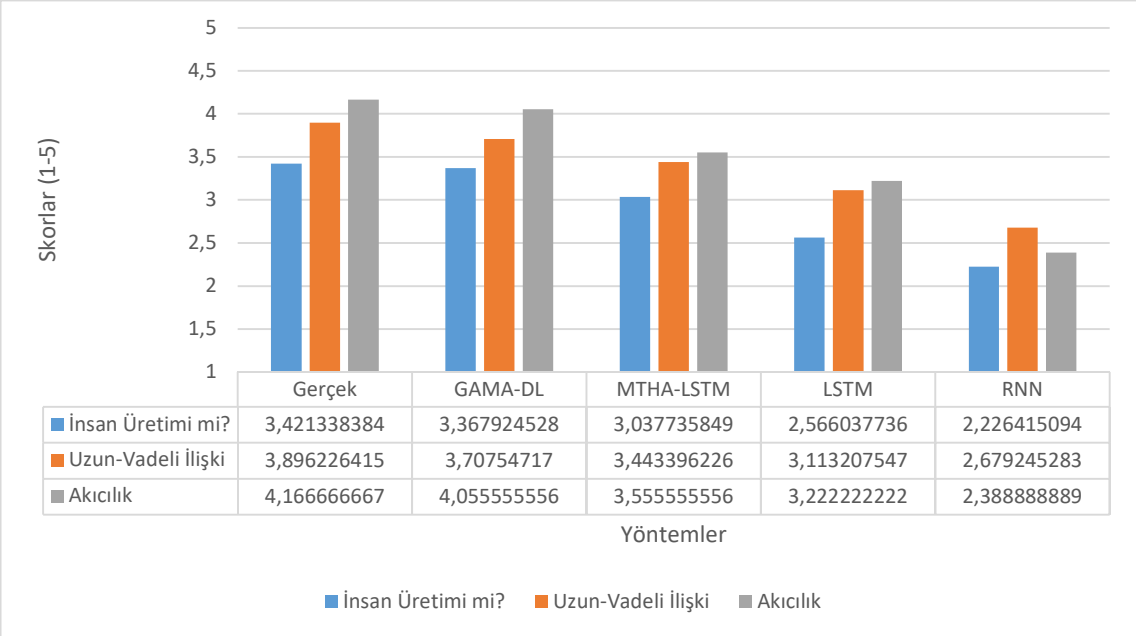
Şekil 4.4'te sunulmakta olan kosinüs benzerliği, müzikal ölçü seviyesinde bir benzerlik analizi sunmaktadır. Yöntemler için elde edilen ortalamalar, kosinüs benzerlik matrisinin üst üçgeninin ortalaması alınarak elde edilmiştir. Her yöntem için iki parça ile kosinüs benzerliği hesaplanmış, elde edilen nihai benzerlik değeri ise bu iki parçanın ortalaması olarak alınmıştır. GAMA-DL'in diğer yöntemlere göre gerçek bestelere en yakın ortalamayı sunması, ölçüler arası ilişkileri yakalamada oldukça başarılı olduğunu ortaya çıkarmaktadır.



Şekil 4.4: Kosinüs benzerlik matrisinde ölçüler arası benzerlik ortalamaları

4.4.2. Nitel Ölçümler

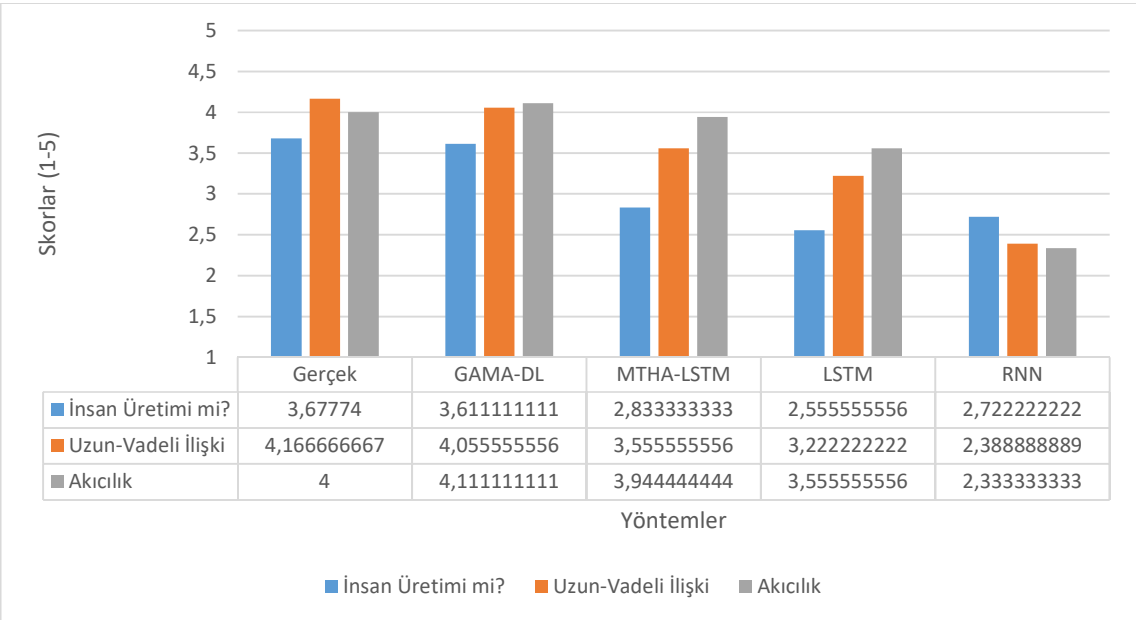
Nitel ölçümler kapsamında GAMA-DL kullanıcı araştırması gerçekleştirilmiştir. Kullanıcı araştırmasında, bir hafta içerisinde 53 katılımcının değerlendirmeleri alınmıştır. Katılımcıların %83'ü kendilerini sıradan dinleyici olarak ifade ederken, %17'si teorik veya mesleki müzikal geçmişi olduğunu ifade etmiştir. Sonuçlar katılımcıların tecrübe seviyelerine göre tümü, tecrübeli ve sıradan dinleyiciler olacak şekilde 3 farklı tabloda incelenmektedir. Şekil 4.5, tüm kullanıcılar için ölçüm sonuçlarını içermektedir. Skorlar 5'li Likert ölçeği kullanıldığından "1-5" arasında dağılıma sahiptir. İlk soruda parçanın "İnsan üretimi" olma durumu sorgulanmış ve "5"e yakın skorlar parçanın "İnsan üretimi", "0"a yakın skorlar ise "Yapay Zekâ üretimi" olduğunu ifade etmektedir. İkinci soruda "Uzun-vadeli İlişki" kurgusu araştırılmış, "5"e yakın skorlar parçanın uzun-vadeli ilişkileri iyi kurguladığını, "0"a yakın skorlar ise parça içerisinde bağlamsal kopukluk olduğunu ifade etmektedir. Son soruda ise "akıcılık" araştırılmış ve parçalarda çalım süresi değişkeninin akıcılığa etkisi bulunmaya çalışılmıştır. Akıcılık için de "5" akıcılığı, "0" ise kesikli veya kopuk ritmik yapıyı ifade etmektedir. GAMA-DL yönteminin kullanıcılar nezdinde tüm metriklerde gerçek müzikal içeriklere en yakın sonuçları ortaya çıkardığı görülmektedir. MTHA-LSTM yöntemi baz yöntemlere göre iyi performans göstermekle birlikte, uzun-vadeli ilişkilerin temsilinde GAMA-DL'in gerisinde kalmaktadır. GAMA-DL'in, nicel istatistiki sonuçlara göre MTHA-LSTM'den daha düşük bir varyans ve daha yüksek ortalama ÖDS ile parça üretebildiği göz önüne alındığında, GAMA-DL'in hem baz yöntemler hem de MTHA-LSTM'e göre çok daha stabil ve özgün içerik üretebildiği anlaşılmaktadır.



	Gerçek	GAMA-DL	MTHA-LSTM	LSTM	RNN
İnsan Üretimi mi?	3,421338384	3,367924528	3,037735849	2,566037736	2,226415094
Uzun-Vadeli İlişki	3,896226415	3,70754717	3,443396226	3,113207547	2,679245283
Akıcılık	4,166666667	4,055555556	3,555555556	3,222222222	2,388888889

Şekil 4.5: GAMA-DL Kullanıcı Çalışması Sonuçları (Tümü)

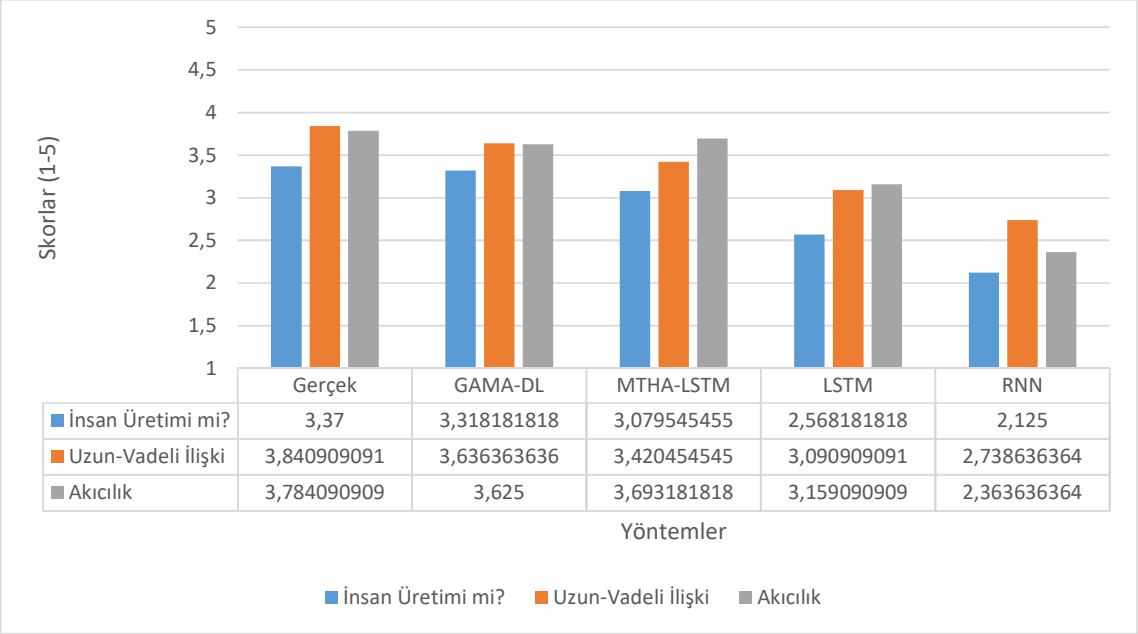
Şekil 4.6, teorik müzik eğitimi almış veya müzikal tecrübesi bulunan katılımcıların skorlarını göstermektedir. Bu klasmandaki sonuçlar genel olarak tüm katılımcıların dağılımıyla uyumluluk göstermektedir. Bununla birlikte, baz modeller RNN ile LSTM arasında tüm metriklerde farkın daha az oluştuğu görülmektedir. Akıcılık noktasında GAMA-DL'in performansını oldukça yüksek bulan katılımcılara göre, modelin çıktıları gerçek parçalar ile de son derece yakın skorlara sahiptir.



	Gerçek	GAMA-DL	MTHA-LSTM	LSTM	RNN
İnsan Üretimi mi?	3,67774	3,611111111	2,833333333	2,555555556	2,722222222
Uzun-Vadeli İlişki	4,166666667	4,055555556	3,555555556	3,222222222	2,388888889
Akıcılık	4	4,111111111	3,944444444	3,555555556	2,333333333

Şekil 4.6: GAMA-DL Kullanıcı Çalışması Sonuçları (Tecrübeli)

Şekil 4.7 müzikal eğitimi veya profesyonel geçmişi bulunmayan katılımcıların değerlendirmelerini içermektedir. Bu kategoride, GAMA-DL'in genel dağılıma göre gerçek parçalardan daha iyi skor aldığı görülmektedir. Bununla birlikte akıcılık kategorisinde en iyi skorları ise MTHA-LSTM yönteminin aldığı görülmektedir.



	Gerçek	GAMA-DL	MTHA-LSTM	LSTM	RNN
İnsan Üretimi mi?	3,37	3,318181818	3,079545455	2,568181818	2,125
Uzun-Vadeli İlişki	3,840909091	3,636363636	3,420454545	3,090909091	2,738636364
Akıcılık	3,784090909	3,625	3,693181818	3,159090909	2,363636364

Şekil 4.7: GAMA-DL Kullanıcı Çalışması Sonuçları (Sıradan Dinleyici)

4.5. Hafıza Karmaşıklığı Analizi

Transformer tabanlı yöntemler, sıralı girdi dizisindeki her bir değişkenin bir diğeri ile ilişkisini sayısal bir değer aracılığı ile doğrudan temsil etmekte ve bu değişkenlerin tamamını matrislerde saklamaktadır. Sıralı girdi uzunluğu n için bu matrislerdeki değişken sayısı n^2 olarak ölçülmektedir. Önerilen MTHA-LSTM ve GAMA-DL yöntemlerinin her ikisi de Transformer tabanlı yöntemlerdir. Bununla birlikte MTHA-LSTM melodi üretiminde tüm girdiyi temsil etmeye çalışırken, GAMA-DL yönteminin nota üretici modelleri bir müzikal ölçü uzunluğunca notayı temsil etmektedirler. Bu sayede nihai müzikal içerik için girdi uzunluğunda yaşanan artışın hafıza karmaşıklığına etkisi sınırlandırılmıştır. Tablo 4.5, MTHA-LSTM ve GAMA-DL için farklı girdi uzunluklarında hafıza karmaşıklığı karşılaştırmasını sunmaktadır. Girdi uzunlukları, sonuçların yorumlanmasını kolaylaştırmak amacıyla doğrusal artış gösterecek şekilde belirlenmiştir.

DÖ yöntemlerinin hafıza karmaşıklığını belirlerken modellerin parametre sayıları dikkate alınmıştır. Modellerin parametre sayıları ile değişkenleri temsil etmekte kullanılan veri tipinin hafızada kapladığı alan çarpılarak hafıza gereksinimi hesaplanmaktadır. MTHA-LSTM ve GAMA-DL için her iki modelde de kullanılan tüm değişkenlerin veri tipi "float32" kullanılmıştır. Bu sebeple tabloda sadece parametre sayıları verilmiştir. GAMA-DL mimarisi müzikal ölçü üreteci, koşulsuz nota üreteci ve müzikal ölçü koşullu nota üreteci olmak üzere üç farklı DÖ yönteminin birlikte kullanımı ile çalıştığından, elde edilen parametre sayısı bu üç yöntemin parametre sayıları toplamını göstermektedir.

Tablo 4.5: Yöntemlerin parametre sayıları

Yöntem	64	128	256	512
GAMA-DL	518K	554K	540K	587K
MTHA-LSTM	458K	1.28M	4.9M	19.3M
LSTM	98K	290K	950K	3.48M
RNN	42K	102K	293K	972K

GAMA-DL yönteminin nota seviyesindeki bileşenleri olan koşulsuz nota üretici ve müzikal ölçü koşullu nota üretici modellerin görevi bir müzikal ölçü uzunluğunda içerik üretmekle sınırlıdır. Böylece girdi vektör uzunluğu ve dahi modelin hafıza karmaşıklığı sabit kalmaktadır. GAMA-DL yöntemi en yüksek öğrenme başarımını ortaya koymakla birlikte, 128 notadan daha uzun müzikal içerikler göz önüne alındığında MTHA-LSTM'e göre daha az parametre kullanarak müzikal içerikleri modelleyebilme kapasitesine sahiptir.

5. TARTIŞMA

Tez araştırmasında geliştirilen yöntemlerin literatürdeki konumunu daha iyi belirleyebilmek amacıyla, yöntem performansları literatürdeki çalışmalar ile nicel ve kullanıcı ölçüm değerlendirmeleri bağlamında karşılaştırılmıştır (Tablo 5.1). Nicel ölçümlerde kategorik çapraz entropi ve MSE kayıp fonksiyonlarının yüksek oranlarda tercih edildiği, dinleme testlerinde ise Likert ölçeğinin yaygın kullanımı görülmektedir. Kategorik çapraz entropi metriği, çok değişkenli çıktı uzaylarında hata skorlarının elde edilmesinde kullanılan popüler bir yöntem olup, tez çalışmasında da kullanılmıştır. Nitel araştırmalarda ise melodik yapının sınanması en popüler yaklaşım olarak öne çıkmaktadır. Bu kapsamda da 5 ölçek büyüklüğünde Likert ölçeği, en sık kullanılan ölçüm yöntemi olarak göze çarpmaktadır. MTHA-LSTM yönteminin nicel ve nitel deney sonuçları, yöntemin literatürdeki mevcut çözümlere göre son derece rekabetçi performans sunduğunu ortaya koymaktadır. GAMA-DL yöntemi ise MTHA-LSTM'in elde ettiği performansın daha da üzerine çıkarak son derece güçlü bir performans ortaya koymuştur. Literatürdeki araştırmaların karşılaştırmalarından çıkarılabilecek bir diğer sonuç ise nicel ve nitel skorların birbirleri ile paralel etkileşim gösterdiğidir. Bu da nicel ve nitel araştırmaların birbirlerini tamamlayıcı ve doğrulayıcı olduğunu göstermektedir.

Tablo 5.1: GAMA-DL ve MTHA-LSTM yöntemlerinin literatürdeki çalışmalar ile karşılaştırması

Yöntemler	Nicel Metrikler	Nicel Skorlar	Nitel Metrikler	Nitel Skorlar
Roberts ve ark. [9]	NLL	1,84	Müzikalite [galibiyet sayısı, 0-60]	>50
Dai ve ark. [21]	Kategorik çapraz entropi	3,20E-03	Melodik yapı [1-5]	>3,00
Liang ve ark. [40]	Boş müzikal ölçü sayısı	6,71	Melodik yapı [1-5]	3,38
Liang ve ark. [41]	Kategorik çapraz entropi	<0,01	Melodik yapı [1-5]	3,68
Vaswani ve ark. [42]	Kategorik çapraz entropi	0,26	Melodik yapı [galibiyet sayısı, 0-128]	>70
Cuthbert ve Ariza [46]	-	-	Kalite, [1-10]	7
Zhang [53]	Benzerlik / Özgün nota	-	Müzikal yapı [1-5]	3,00
Huang ve Wu [56]	MSE	0.00196	-	-
Mangal ve ark. [57]	MSE	0.22	-	-
MTHA-LSTM	Kategorik çapraz entropi	9,16E-03	Melodik yapı [1-5]	3,15
GAMA-DL	Kategorik çapraz entropi	7,24E-07	Melodik yapı [1-5]	3,7

6. SONUÇ ve ÖNERİLER

Bu tezde, çok-sesli müzikal içerik üretimi problemi, tek-sesli melodik içerik üretimi ve tek-sesli eşlik üretimi olmak üzere iki alt-problem nezdinde incelenmiştir. Çalışmada müzikal içerik üretiminde yenilikçi ve çok-seviyeli hiyerarşik bir mimari ile çok-sesli müzikal içerik üretimi probleminin çözümü amaçlanmıştır. Elde edilen nihai eser, eşlik ile zenginleştirilmiş melodik bir ezgiyi ifade eden çok-sesli müzikal bir içeriktir.

Tek-sesli melodi üretimi probleminin çözümünde iki özgün yöntem önerilmiştir. Çok-başlıklı dikkat ağları ile uzun-kısa vadeli hafıza (MTHA-LSTM) ağlarının birlikte kullanımı ile uzun-vadeli ilişkileri müzikal değişkenler üzerinde güçlü bir şekilde temsil edebilen ÜYZ yöntemi geliştirilmiştir. İkinci yöntemde, müzikal ölçü ve nota seviyesinde birbirleri ile entegre çalışan üretici ağların melodi üretimi probleminde kullanıldığı, ölçü-koşullu müzik-dikkat ÜYZ yöntemi (GAMA-DL) geliştirilmiştir. GAMA-DL yöntemi, müzikal ölçü seviyesinde çizge temsili, nota seviyesinde ise sıralı sembolik gösterimden yararlanmıştır. Geliştirilen mimari ile referans yöntemlerden daha güçlü bir uzun-vadeli temsil sağlanmış ve nota seviyesinde dikkat ağlarına sağlanan girdi vektörünü sınırlandırılarak dikkat ağlarının hafıza karmaşıklığı sorunu ortadan kaldırılmıştır.

Elde edilen sonuçların geçerliliği ve güvenirliği, MAESTRO klasik tonal batı müziği veri setinde gerçekleştirilen nicel ve nitel deneyler ile desteklenerek incelenmiştir. KPM veriseti, DÖ yöntemlerinin hiper-parametre optimizasyonunda kullanılmıştır. Yöntemlerin müzikal kalite ve uzun-vadeli ilişki temsil kabiliyeti açısından yapılan istatistiksel analizler, geliştirilen yöntemlerin literatürdeki mevcut yöntemlere göre daha etkin olduklarını göstermektedir.

Transformer mimarilerin yüksek hesaplama karmaşıklığı oluşturması sebebiyle analizler tek-enstrümanlı müzikal içerik uzayıyla sınırlandırılmıştır. Daha büyük veri setleri ve donanımsal kapasitenin kullanımı değerlendirilerek çok-enstrümanlı müzikal içerik üretimi problemlerinde çizge gösteriminin etkileri incelenebilir. Çizge gösteriminin müzikal gösterim olarak sembolik ve sinyal gösterimlere göre karmaşık ve hiyerarşik özellikleri temsilde önemli avantajları bulunmaktadır. Nota ve müzikal ölçü seviyesinde tez çalışmasında önerilenden daha farklı veya daha fazla ilişkinin incelenmesi yine cevabı aranabilecek araştırma soruları arasında

bulunmaktadır. Çizge Sinir Ağları'nın, çizgelerde gerçekleştirilebilecek analizlerde kullanılabilecek faydalı bir alternatif yöntem olarak yakın tarihli çalışmalarda öne çıktığı görülmektedir. Bu ağların kullanıldığı yöntemler ile müzikal içerik üretimi sınanabilir. ÜYZ yöntemleri ile müzikal içerik üretiminde kullanılan nicel ölçüm metriklerinin olgunlaşmaya ihtiyacı bulunmaktadır. Üzerinde konsensüs sağlanmış ve müzikal açıdan yeterliliği ifade eden nicel metriklerin geliştirilmesi önemli bir araştırma sorusudur. Tez çalışmasında literatüre paralel olarak DDİ ve MIR alanlarındaki ortak nicel metrikler kullanılmıştır. Müzik teorisi ele alınarak, müzikal bilginin temsilinde hiyerarşik ve çok katmanlı yapıyı daha verimli sunabilecek gösterimler elde edilebilecektir.

KAYNAKLAR

[1] Burgess, R. J., 2014, *The history of music production*, Oxford University Press, Oxford.

[2] Downie, J. S., 2003, Music information retrieval, *Annual review of information science and technology*, 37 (1), 295-340.

[3] Carlson, K., Kopalle, P. K., Riddell, A., Rockmore, D., Vana, P., 2023, Complementing human effort in online reviews: A deep learning approach to automatic content generation and review synthesis, *International Journal of Research in Marketing*, 40 (1), 54-74.

[4] Tang, C. and Zhang, J., 2022, An intelligent deep learning-enabled recommendation algorithm for teaching music students, *Soft Computing*, 26 (20), 10591–10598.

[5] Koh, E. S., Dubnov, S., Wright, D., 2018, Rethinking recurrent latent variable model for music composition, *International Workshop on Multimedia Signal Processing*, IEEE, 1–6.

[6] Cao, J., Fiaidhi, J., Qi, M., 2020, A Review of Automatic Music Generation Based on Performance RNN, *TechRxiv*.

[7] Briot, J. P., Hadjeres, G., Pachet, F. D., 2020, *Deep learning techniques for music generation*, Springer, İsviçre.

[8] Bengio, Y., Simard, P., Frasconi, P., 1994, Learning long-term dependencies with gradient descent is difficult, *IEEE transactions on neural networks*, 5(2), 157-166.

[9] Roberts, A., Engel, J., Raffel, C., Hawthorne, C., Eck, D., 2018, A hierarchical latent vector model for learning long-term structure in music, *International conference on machine learning*, PMLR, 4364-4373.

[10] Loyola, F., 2021, *Polyphonic music generation using neural networks*, Thesis (Master's), Universitat Politècnica de Catalunya.

[11] Wang, Z., Wang, D., Zhang, Y., Xia, G., 2020, *Learning interpretable representation for controllable polyphonic music generation*, http://arxiv.org/abs/2008.07122, [Ziyaret Tarihi: 1 Ocak 2024].

[12] Ji, S., Luo, J., Yang, X., 2020, *A comprehensive survey on deep music generation: Multi-level representations, algorithms, evaluations, and future directions*, http://arxiv.org/abs/2011.06801, [Ziyaret Tarihi: 1 Ocak 2024].

[13] Qin, Y., Xie, H., Ding, S., Tan, B., Li, Y., Zhao, B., Ye, M., 2023, Bar transformer: a hierarchical model for learning long-term structure and generating impressive pop music, *Applied Intelligence*, 53 (9), 10130-10148.

[14] Huang, W., Yu, Y., Xu, H., Su, Z., Wu, Y., 2023, Hyperbolic music transformer for structured music generation, *IEEE Access*, 11, 26893-26905.

[15] Gibson, J., 2013, *The MIDI Standard: Introduction to MIDI and Computer Music.*, https://cecm.indiana.edu/361/midi.html, [Ziyaret tarihi: 1 Ocak 2024].

[16] Rothstein, J., 1995, *MIDI: A comprehensive introduction*, (7), AR Editions, USA.

[17] Zhang, H., Karystinaios, E., Dixon, S., Widmer, G., Cancino-Chacón, C. E., 2023, *Symbolic Music Representations for Classification Tasks: A Systematic Evaluation*, http://arxiv.org/abs/2309.02567, [Ziyaret Tarihi: 1 Ocak 2024].

[18] Jeong, D., Kwon, T., Kim, Y., Nam, J., 2019, Graph neural network for music score data and modeling expressive piano performance, *International conference on machine learning*, PMLR, 3060-3070.

[19] Karystinaios, E., Widmer, G., 2022, Cadence Detection in Symbolic Classical Music using Graph Neural Networks, *Proceedings of the 22nd International Society of Music Information Retrieval Conference*, ISMIR, 1-5.

[20] Guo, C., Liu, X., 2015, Automatic feature generation on heterogeneous graph for music recommendation, *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, 807-810.

[21] Dai, S., Jin, Z., Gomes, C., Dannenberg, R. B., 2021, *Controllable deep melody generation via hierarchical music structure representation*, http://arxiv.org/abs/2109.00663, [Ziyaret Tarihi: 1 Ocak 2024].

[22] Capone, V., Iannuzzo, G., Camastra, F., 2023, Deep Learning for Time Series Forecasting: Advances and Open Problems, *Information*, 14 (11), 598.

[23] Ozcan, A., Catal, C., Kasif, A., 2021, Energy load forecasting using a dual-stage attention-based recurrent neural network, *Sensors*, 21 (21), 7115.

[24] Wang, X., Ma, J. J., Wang, S., Bi, D. W., 2007, Time series forecasting for energy-efficient organization of wireless sensor networks, *Sensors*, 7 (9), 1766-1792.

[25] Karevan, Z., Suykens, J. A., 2020, Transductive LSTM for time-series prediction: An application to weather forecasting, *Neural Networks*, 125, 1-9.

[26] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014, Learning phrase representations using RNN encoder-decoder for statistical machine translation, http://arxiv.org/abs/1406.1078, [Ziyaret Tarihi: 1 Ocak 2024].

[27] Hochreiter, S., Schmidhuber, J., 1997, Long short-term memory, *Neural computation*, 9 (8), 1735-1780.

[28] Mei, X., Liu, X., Sun, J., Plumbley, M. D., Wang, W., 2022, Diverse audio captioning via adversarial training, International Conference on Acoustics, Speech and Signal Processing, IEEE, 8882-8886.

[29] Hennig, J. A., Umakantha, A., Williamson, R. C., 2017, *A classifying variational autoencoder with application to polyphonic music generation*, http://arxiv.org/abs/1711.07050, [Ziyaret Tarihi: 1 Ocak 2024].

[30] Chung, J., Gulcehre, C., Cho, K., Bengio, Y., 2014, *Empirical evaluation of gated recurrent neural networks on sequence modeling*, http://arxiv.org/abs/arXiv:1412.3555, [Ziyaret Tarihi: 1 Ocak 2024].

[31] Wu, J., Liu, X., Hu, X., Zhu, J., 2020, PopMNet: Generating structured pop music melodies using neural networks, *Artificial Intelligence*, 286, 103303.

[32] Kingma, D. P., Welling, M., 2013, *Auto-encoding variational bayes*, http://arxiv.org/abs/1312.6114, [Ziyaret Tarihi: 1 Ocak 2024].

[33] Dieleman, S., Schrauwen, B., 2014, End-to-end learning for music audio, *International Conference on Acoustics, Speech and Signal Processing*, IEEE, 6964-6968.

[34] Choi, K., Fazekas, G., Sandler, M., Cho, K., 2017, Convolutional recurrent neural networks for music classification, *International conference on acoustics, speech and signal processing*, IEEE, 2392-2396.

[35] Karpathy, A., Johnson, J., Fei-Fei, L., 2015, Visualizing and understanding recurrent networks, http://arxiv.org/abs/1506.02078, [Ziyaret Tarihi: 1 Ocak 2024].

[36] Boulanger-Lewandowski, N., Bengio, Y., Vincent, P., 2012, Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription, http://arxiv.org/abs/1206.6392, [Ziyaret Tarihi: 1 Ocak 2024].

[37] Zixun, G., Makris, D., Herremans, D., 2021, Hierarchical recurrent neural networks for conditional melody generation with long-term structure, *International Joint Conference on Neural Networks*, IEEE, 1-8.

[38] Eck, D., Schmidhuber, J., 2002, Finding temporal structure in music: Blues improvisation with LSTM recurrent networks, *Proceedings of the 12th IEEE workshop on neural networks for signal processing*, IEEE, 747-756.

[39] Pascanu, R., Mikolov, T., Bengio, Y., 2013, On the difficulty of training recurrent neural networks, *International conference on machine learning*, PMLR, 1310-1318.

[40] Liang, X., Wu, J., Yin, Y., 2019, *MIDI-sandwich: multi-model multi-task hierarchical conditional VAE-GAN networks for symbolic single-track music generation*, http://arxiv.org/abs/1907.01607, [Ziyaret Tarihi: 1 Ocak 2024].

[41] Liang, X., Wu, J., Cao, J., 2019, *MIDI-Sandwich2: RNN-based Hierarchical Multi-modal Fusion Generation VAE networks for multi-track symbolic music generation*, http://arxiv.org/abs/1909.03522, [Ziyaret Tarihi: 1 Ocak 2024].

[42] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I., 2017, Attention is all you need, *Advances in Neural Information Processing Systems*, 30.

[43] Huang, C. Z. A., Vaswani, A., Uszkoreit, J., Shazeer, N., Simon, I., Hawthorne, C., … Eck, D., 2018, *Music transformer*, http://arxiv.org/abs/1809.04281, [Ziyaret Tarihi: 1 Ocak 2024].

[44] Krueger, B., 2023, *Classical Piano Midi Dataset*, http://www.piano-midi.de/midi_files.html, [Ziyaret Tarihi: 1 Ocak 2024].

[45] Hawthorne, C., Stasyuk, A., Roberts, A., Simon, I., Huang, C.-Z. A., Dieleman, S., … Eck, D., 2018, *Enabling factorized piano music modeling and generation with the MAESTRO dataset*, http://arxiv.org/abs/1810.12247, [Ziyaret Tarihi: 1 Ocak 2024].

[46] Cuthbert, M. S., Ariza, C., 2010, music21: A toolkit for computer-aided musicology and symbolic music data, *Proceedings of the 11th International Society for Music Information Retrieval Conference*, ISMIR.

[47] Gu, J., Bradbury, J., Xiong, C., Li, V. O., Socher, R., 2017, Non-autoregressive neural machine translation, http://arxiv.org/abs/1711.02281, [Ziyaret Tarihi: 1 Ocak 2024].

[48] Elman, J. L., 1990, Finding structure in time, *Cognitive science*, 14(2), 179-211.

[49] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014, Dropout: a simple way to prevent neural networks from overfitting, *The journal of machine learning research*, 15 (1), 1929-1958.

[50] Cosenza, E., Valenti, A., Bacciu, D., 2023, Graph-based polyphonic multitrack music generation, http://arxiv.org/abs/2307.14928, [Ziyaret Tarihi: 1 Ocak 2024].

[51] Zou, Y., Zou, P., Zhao, Y., Zhang, K., Zhang, R., Wang, X., 2022, Melons: generating melody with long-term structure using transformers and structure graph, *International Conference on Acoustics, Speech and Signal Processing*, IEEE, 191-195.

[52] Go, S., Lee, K., 2023, *Combinatorial music generation model with song structure graph analysis*, http://arxiv.org/abs/2312.15400, [Ziyaret Tarihi: 1 Ocak 2024].

[53] Zhang, N., 2020, Learning adversarial transformer for symbolic music generation. *IEEE transactions on neural networks and learning systems*, 34(4), 1754-1763.

[54] Yamshchikov, I. P., Tikhonov, A., 2020, Music generation with variational recurrent autoencoder supported by history, *SN Applied Sciences*, 2 (12), 1937.

[55] Joshi, A., Kale, S., Chandel, S., Pal, D. K., 2015, Likert scale: Explored and explained, *British Journal of Applied Science & Technology*, 7(4), 396.

[56] Huang, A., Wu, R., 2016, *Deep learning for music*, http://arxiv.org/abs/1606.04930, [Ziyaret Tarihi: 1 Ocak 2024].

[57] Mangal, S., Modak, R., Joshi, P., 2019, *Lstm based music generation system*, http://arxiv.org/abs/1908.01080, [Ziyaret Tarihi: 1 Ocak 2024].

EKLER

İNTİHAL RAPORU İLK SAYFASI

Ahmet Kaşif

ORİJİNALLİK RAPORU

%7	%6	%3	%5
BENZERLİK ENDEKSİ	İNTERNET KAYNAKLARI	YAYINLAR	ÖĞRENCİ ÖDEVLERİ

BİRİNCİL KAYNAKLAR

1	Submitted to The Scientific & Technological Research Council of Turkey (TUBITAK) Öğrenci Ödevi	%3
2	acikbilim.yok.gov.tr İnternet Kaynağı	%1
3	research.bau.edu.tr İnternet Kaynağı	<%1
4	link.springer.com İnternet Kaynağı	<%1
5	www-staff.it.uts.edu.au İnternet Kaynağı	<%1
6	Submitted to Indiana Wesleyan University Öğrenci Ödevi	<%1
7	ceng.btu.edu.tr İnternet Kaynağı	<%1
8	jise.btu.edu.tr İnternet Kaynağı	<%1
9	Submitted to University of Pretoria Öğrenci Ödevi	<%1

ETİK KURUL İZİN YAZISI

Uyarı: Canlı denekler üzerinde yapılan tüm araştırmalar için Etik Kurul Belgesi alınması zorunludur.

☐ **Etik Kurul izni gerekmektedir.**

☒ **Etik Kurul izni gerekmemektedir.**

Ahmet KAŞİF
(İmza)

KURUM İZNİ YAZILARI

Uyarı: Canlı ve cansız deneklerle yapılan tüm çalışmalar için kurum izin belgelerinin eklenmesi zorunludur. Gizlilik ve mahremiyet içeren durumlarda kurum adı kapatılmalıdır.

☐ **Kurum izni gerekmektedir.**

☒ **Kurum izni gerekmemektedir.**

Ahmet KAŞİF
(İmza)