



دانشگاه صنعتی امیرکبیر  
(پلی تکنیک تهران)  
دانشکده مهندسی کامپیوتر

پایان نامه کارشناسی

# ارزیابی ترکیب شبکه‌های عصبی گرافی و مدل پوینت‌نت برای پردازش داده‌های گرافی و ابرنقاط

نگارش  
محمدچوپان

استاد راهنما  
دکتر مهدی جوانمردی

اردیبهشت ۱۴۰۳

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه صنعتی امیرکبیر  
(پلی تکنیک تهران)  
دانشکده مهندسی کامپیوتر

پایان نامه کارشناسی

# ارزیابی ترکیب شبکه‌های عصبی گرافی و مدل پوینت‌نت برای پردازش داده‌های گرافی و ابرنقاط

نگارش

محمد چوپان

استاد راهنما

دکتر مهدی جوانمردی

اردیبهشت ۱۴۰۳

# صفحه فرم ارزیابی و تصویب پایان نامه - فرم تأیید اعضاء کمیته دفاع

در این صفحه فرم دفاع یا تأیید و تصویب پایان نامه موسوم به فرم کمیته دفاع - موجود در پرونده آموزشی - را قرار دهید.

## نکات مهم:

- نگارش پایان نامه/رساله باید به **زبان فارسی** و بر اساس آخرین نسخه دستورالعمل و راهنمای تدوین پایان نامه های دانشگاه صنعتی امیرکبیر باشد.(دستورالعمل و راهنمای حاضر)
- رنگ جلد پایان نامه/رساله چاپی کارشناسی، کارشناسی ارشد و دکترا باید به ترتیب مشکی، طوسی و سفید رنگ باشد.
- چاپ و صحافی پایان نامه/رساله بصورت **پشت و رو(دورو)** بلامانع است و انجام آن توصیه می شود.



دانشگاه صنعتی امیرکبیر  
(پلی تکنیک تهران)

به نام خدا

## تعهدنامه اصالت اثر

تاریخ: اردیبهشت ۱۴۰۳

اینجانب **محمد چوپان** متعهد می‌شوم که مطالب مندرج در این پایان‌نامه حاصل کار پژوهشی اینجانب تحت نظارت و راهنمایی اساتید دانشگاه صنعتی امیرکبیر بوده و به دستاوردهای دیگران که در این پژوهش از آنها استفاده شده است مطابق مقررات و روال متعارف ارجاع و در فهرست منابع و مآخذ ذکر گردیده است. این پایان‌نامه قبلاً برای احراز هیچ مدرک هم‌سطح یا بالاتر ارائه نگردیده است. در صورت اثبات تخلف در هر زمان، مدرک تحصیلی صادر شده توسط دانشگاه از درجه اعتبار ساقط بوده و دانشگاه حق پیگیری قانونی خواهد داشت.

کلیه نتایج و حقوق حاصل از این پایان‌نامه متعلق به دانشگاه صنعتی امیرکبیر می‌باشد. هرگونه استفاده از نتایج علمی و عملی، واگذاری اطلاعات به دیگران یا چاپ و تکثیر، نسخه‌برداری، ترجمه و اقتباس از این پایان‌نامه بدون موافقت کتبی دانشگاه صنعتی امیرکبیر ممنوع است. نقل مطالب با ذکر مآخذ بلامانع است.

**محمد چوپان**

امضا

تقدیم به پدر، مادر و خواهر عزیزم که در سختی ها و دشواری های زندگی، همواره یاری  
دلسوز و پشتیبانی محکم و مطمئن برایم بوده اند.

# سپاس‌گزاری

زندگی دفتري از خاطره هاست، يك نفر در دل شب، يك نفر در دل خاك، يك نفر همدم خوشبختي هاست، يك نفر همسفر سختي هاست، چشم تا باز كنيم، عمرمان مي‌گذرد ما همه رهگذريم، آنچه باقيست فقط خوبيهاست.

تشكر مي‌كنم از تمامي عزيزاني كه در تمامي مراحل زندگي همراه من بوده اند. و همچنين از استاد گرامي جناب آقاي دكتر مهدي جوانمردي كه درانتخاب و پيشبرد اين پروژه به عنوان استاد پروژه، كمكهاي فراواني به اين جانب داشتند، كمال تشكر را دارم.

محمد چوپان  
اردیبهشت ۱۴۰۳

## چکیده

این پایان نامه به اکتشاف یادگیری عمیق و داده‌های ابرنقاط و گرافی می‌پردازد و نحوه عملکرد معماری‌های متفاوت در دسته‌بندی و پردازش این دو داده را ارزیابی می‌کند. ابتدا ما به بررسی خودروهای خودران و اهمیت آن‌ها در زندگی روزمره می‌پردازیم. سپس معرفی داده‌های ابرنقاط<sup>۱</sup> و گرافی و کاربرد آن‌ها در زمینه خودروهای خودران را بررسی می‌کنیم و مجموع داده‌های استفاده شده در این پروژه را معرفی می‌کنیم. پس از آن بررسی معماری‌های تجمیع گراف خودتوجه<sup>۲</sup> و پوینت نت<sup>۳</sup> می‌پردازیم. در قدم بعدی معرفی معماری‌های حاصل از ترکیب این دو و نتایج حاصل از دو معماری به صورت مجزا و در نهایت نتایج به دست آمده از ترکیب این دو معماری را بررسی می‌کنیم. در قدم آخر نیز تاثیر هر یک از نقاط داده‌های خود را به صورت جدا بررسی کرده و بهبودهای داده شده در عملکرد معماری معرفی شده را با استفاده از نتایج شهودی بررسی خواهیم کرد.

## واژه‌های کلیدی:

یادگیری عمیق، ابرنقاط، معماری تجمیع خودتوجه، پوینت نت، خودروهای خودران

---

<sup>1</sup>Point Cloud Data

<sup>2</sup>Self Attention Graph Pooling

<sup>3</sup>Point NeT



# فهرست مطالب

صفحه

عنوان

۱	.....	۱ مقدمه
۳	.....	۲ موسیقی
۴	.....	۱-۲ تاثیر آن در زندگی
۴	.....	۲-۲ در صنایع مختلف
۴	.....	۱-۲-۲ تبلیغات
۵	.....	۲-۲-۲ فیلم سازی
۶	.....	۳-۲-۲ درمانی
۶	.....	۳-۲ خلاصه
۷	.....	۳ هوش مصنوعی
۸	.....	۱-۳ نگاهی بر علم هوش مصنوعی
۸	.....	۱-۱-۳ یادگیری ماشین
۹	.....	۲-۱-۳ یادگیری عمیق
۱۰	.....	۲-۳ خلاصه
۱۲	.....	۴ نحوه تبدیل موسیقی به زبان رایانه
۱۳	.....	۱-۴ مقاطاعات پس زمینه موسیقی
۱۴	.....	۱-۱-۴ نوع داده
۱۵	.....	۲-۱-۴ شبکه های عصبی
۱۶	.....	۲-۴ معماری پیشنهادی
۱۶	.....	۱-۲-۴ تبدیل داده ها به نوع آهنگ
۱۷	.....	۲-۲-۴ رمزگذاری
۱۸	.....	۳-۲-۴ افزایش داده ها
۱۹	.....	۴-۲-۴ گام های زمانی
۲۰	.....	۳-۴ خلاصه

۲۲	۵ نحوه ساخت موسیقی
۲۳	۱-۵ شبکه عصبی
۲۳	۲-۵ تولید موسیقی
۲۴	۳-۵ داده ها و ابزار ها
۲۴	۱-۳-۵ مجموعه داده ها
۲۵	۲-۳-۵ ابزار های استفاده شده
۲۷	۳-۳-۵ شبکه عصبی
۲۷	۴-۵ نتایج
۲۹	۶ نتیجه گیری و پیشنهادات
۳۰	۱-۶ نتیجه گیری
۳۱	۲-۶ پیشنهادات
۳۲	کتابنامه

## فهرست تصاویر

شکل	صفحه
۱-۴ کیبورد پیانو . . . . .	۱۳
۲-۴ نمونه یک تصویر از صفحه برنامه . . . . .	۱۴
۳-۴ نمونه یک پیام . . . . .	۱۵
۴-۴ نقشه راه پیشنهادی . . . . .	۱۶
۵-۴ فرایند تبدیل نوع آهنگ . . . . .	۱۷
۶-۴ روش کد گذاری . . . . .	۱۹
۷-۴ گذر از نوت سی ماژور به اف ماژور . . . . .	۱۹
۸-۴ نمونه یک مقاله در گوگل اسکولار . . . . .	۲۰
۹-۴ نمونه یک مقاله در گوگل اسکولار . . . . .	۲۱
۱-۵ آموزش شبکه عصبی . . . . .	۲۳
۲-۵ روش تولید موسیقی . . . . .	۲۴
۳-۵ شماره کلید های مجموعه داده ها محور افقی شماره کلید و محور عمودی فرکانس عادی شده . . . . .	۲۵
۴-۵ محور افقی شماره کلید پیانو و محور عمودی فرکانس عادی نشده . . . . .	۲۶
۵-۵ معماری شبکه عصبی . . . . .	۲۸

صفحه	فهرست جداول	جدول
------	-------------	------

# فصل اول

## مقدمه

موسیقی در زندگی انسان نقش حیاتی را دارا است. علاوه بر زندگی روزمره انسان در صنعت هایی مانند فیلم سازی، تبلیغات، درمان و ... کاربر های متفاوتی دارد. حال ساخت یک موسیقی دل نشین و کاربردی چالشی است که امروزه اکثر هنرمندان در این زمینه با آن درگیرند. در این پژوهش نگاهی می اندازیم بر بررسی و ساخت موسیقی با استفاده از هوش مصنوعی، بررسی چالش ها و روش های متفاوت انجام این کار و در انتها بر تاثیر موسیقی های ساخته شده بر روی صنایع متفاوت نگاهی می اندازیم. در فصل ابتدایی به بررسی موسیقی و تاثیر آن بر روی زندگی روزمره می پردازیم. سپس در فصل دوم به بررسی هوش مصنوعی و الگوریتم ها متفاوت آن می پردازیم. در فصل سوم نیز نگاهی میکنیم به نحوه تبدیل موسیقی به زبان رایانه و برعکس آن، سپس نحوه ساخت موسیقی با استفاده از هوش مصنوعی را بررسی میکنیم. در فصل چهارم نگاهی می اندازیم بر موسیقی های ساخته شده با استفاده از هوش مصنوعی ، در آخر نیز نتیجه از مقایسه موسیقی های ساخته شده با این روش و موسیقی های نواخته شده و چشم اندازی از آینده این موسیقیها می آوریم.

فصل دوم

موسیقی

## ۱-۲ تاثیر آن در زندگی

موسیقی در تمامی بخش‌های زندگی ما وجود دارد. اعم از زندگی روزمره، کار، حتی هنگام دیدن یک فیلم و یا خواب می‌توان تاثیر آن را مشاهده کرد. بر اساس مشاهدات گوش دادن به موسیقی‌های متفاوت بر روی کاهش درد و یا عمل کرد بهتر مغز انسان تاثیر دارد. در آزمایشی که بر روی کارکنان شرکت‌های مختلف کانادایی انجام شده است، مطالعات نشان می‌دهد که به طور میانگین عملکرد آن‌ها ۱۵ درصد افزایش داشته است. علاوه بر این‌ها گوش دادن به موسیقی بر غذا خوردن هم تاثیر دارد. مطالعات نشان می‌دهد که گوش دادن به موسیقی سریع هنگام غذا خوردن، روند غذا خوردن را سریع‌تر می‌کند. و بالعکس، گوش دادن به موسیقی هنگام فعالیت‌های بدنی موجب می‌شود که آهنگ فعالیت‌های انسان بیشتر شده و فعالیت مورد نظر را سریع‌تر انجام دهد. و یا در مواقعی که برای انسان استرس‌زا است گوش دادن به موسیقی باعث کاهش ۲۶ درصدی اضطراب انسان می‌شود. [۵۱]

## ۲-۲ در صنایع مختلف

موسیقی علاوه بر زندگی روزمره در صنایع مختلفی مانند فیلم سازی، تبلیغات، درمان و دارویی نیز کاربرد دارد. در ادامه به کاربردهای آن در صنایع مختلف می‌پردازیم.

۱. تبلیغات

۲. فیلم سازی

۳. درمانی

### ۱-۲-۲ تبلیغات

آشنایی با موسیقی پس‌زمینه بیشتر بر نگرش‌ها نسبت به تبلیغات، نگرش‌های برند و قصد خرید در شرایط مشارکت کم تاثیر می‌گذارد، در حالی که تناسب محصول با موسیقی پس‌زمینه بیشتر در شرایط مشارکت بالا تاثیرگذار است. این نشان می‌دهد که ویژگی‌های مؤثر موسیقی پس‌زمینه می‌تواند با توجه به مشارکت مصرف‌کننده متفاوت باشد. اگرچه موسیقی در تبلیغات می‌تواند تأثیر بسیار مثبتی بر مصرف‌کننده داشته باشد، اما نگرش نسبت به محصول تجاری و نام تجاری نیز بسیار می‌تواند متفاوت باشد. مشکلات هنگام استفاده از موسیقی همیشه بخشی از مردم هستند که از موسیقی ای که در یک آگهی



بازرگانی پخش می شود، خوششان نمی آید، خواه به خاطر آن باشد که هنرمند، ژانر یا سبک را دوست ندارند. این را نمی توان کنترل کرد زیرا که این یک عامل ذهنی بسته به شخصیت و ویژگی های فردی است. علاوه بر این، موسیقی همچنین می تواند به عنوان یک حواس پرتی و محدودیت برای مصرف کنندگان عمل کند که سعی در پردازش اطلاعات ارائه شده در آگهی دارند. [۶، ۷]

## ۲-۲-۲ فیلم سازی

در این بخش نتایج مطالعه افراد بر روی تاثیر یک موسیقی فیلم بر روی احساسات بیننده را مشاهده میکنیم. چهار جفت عواطف قطبی تعریف شده در مدل پلاچیک به عنوان ویژگی های عاطفی اساسی مورد استفاده قرار گرفتند: شادی-غم، انتظار-تعجب، ترس-خشم و اعتماد-انزجار. در مطالعه اولیه، هشت سکانس فیلم و هشت تم موسیقی به عنوان بهترین نمایندگان هر هشت احساس پلاچیک انتخاب شدند. در آزمایش اصلی، شرکت کنندگان کیفیت های احساسی ترکیب های فیلم-موسیقی را در هشت مقیاس هفت درجه ای قضاوت کردند. نیمی از ترکیب ها همخوان بودند (مثلاً فیلم شاد - موسیقی شاد) و نیمی ناسازگار بودند (مثلاً فیلم شاد - موسیقی غمگین). نتایج نشان داده است که اطلاعات بصری (فیلم) تأثیر بیشتری بر ارزیابی هیجان نسبت به اطلاعات شنیداری (موسیقی) دارد. اثرات مدل سازی پس زمینه موسیقی به کیفیت های احساسی بستگی دارد. در برخی از ترکیب های نامتجانس (شادی-غم) تعدیل هایی در جهت های مورد انتظار به دست آمد (مثلاً موسیقی شاد غم یک فیلم غمگین را کاهش می دهد)، در برخی موارد (خشم-ترس) هیچ اثر مدل سازی به دست نیامد و در برخی موارد (اعتماد)-انزجار، انتظار-غافلگیری) اثرات مدل سازی در جهت غیرمنتظره بود (مثلاً موسیقی قابل اعتماد ارزیابی انزجار از یک فیلم منجر کننده را افزایش داد). این نتایج نشان می دهد که ارزیابی تأثیرات مشترک احساسات به رسانه (فیلم موسیقی را می پوشاند) و کیفیت احساسی (سه نوع اثر مدل سازی) بستگی دارد. دو آزمایش تأثیر موسیقی فیلم را بر اقناع روایت بررسی کردند. در آزمایش اول، شرکت کنندگان یک فیلم کوتاه را با موسیقی متن اصلی آن یا با موسیقی متن بی صدا تماشا کردند. در آزمایش دوم، موسیقی متن موسیقی به فیلمی اضافه شد که در ابتدا بدون موسیقی تولید شده بود. یافته ها حاکی از آن بود که شرکت کنندگان هنگام ارائه موسیقی متن فیلم، انتقال بیشتر به فیلم و توافق بیشتر با باورهای مرتبط با فیلم را گزارش کردند، اما تنها زمانی که موسیقی با لحن عاطفی فیلم مطابقت داشت. در حمایت از مدل حمل و نقل-تصویر<sup>۱</sup> تأثیر موسیقی فیلم بر باورها با انتقال به فیلم واسطه شد. [۸، ۹]

<sup>1</sup>Green and Brock transportation theory

## ۳-۲-۲ درمانی

در این بخش نتایج مطالعه افراد بر روی تأثیر گوش دادن به موسیقی ترجیحی بر ادراک اضطراب و درد بیماران تحت همودیالیز را مشاهده میکنیم. در انجام این آزمایش از دو گروه از افراد استفاده شده است. شصت نفر با تشخیص نارسایی کلیوی مرحله نهایی تحت درمان همودیالیز در این مطالعه شرکت کردند. گوش دادن به موسیقی ترجیحی به عنوان مداخله استفاده شده است. اضطراب و درد در پیش آزمون و پس آزمون اندازه گیری شده است. گروه کنترل در مرحله پس آزمون در اضطراب حالت به طور معنی داری نمره بالاتری نسبت به گروه آزمایش شده کسب کردند و شدت درد بالاتری را به طور معنی داری در مرحله پس آزمون تجربه کردند. این مطالعه نشان داد که گوش دادن به موسیقی در حین کشیدن غده بدخیم سوم نهفته فک پایین باعث سرکوب فعالیت اعصاب سمپاتیک<sup>۲</sup>

در حین برش، برداشتن استخوان و جدا شدن تاج دندان می شود و اضطراب پس از درمان را تسکین می دهد. مطالعات آتی بر روی مکانیسم های درگیر و روش های پیشگیری از شروع حوادث سیستمیک تمرکز خواهند کرد. نمره پیش آزمون نشان داد که اکثر بیماران دارای سطح متوسط (۵/۷۷ درصد) و ۱۵ درصد دارای سطح اضطراب حالت شدید بودند. اندازه گیری مکرر نشان داد که میانگین نمره اضطراب حالت در هر یک از سه اندازه گیری پس آزمون در گروه آزمایش به طور قابل توجهی کمتر از گروه کنترل بود. میانگین تفاوت بین گروه آزمایش و کنترل مقدار معنا داری است. [۱۲۱۰]

## ۳-۲ خلاصه

با توجه به مطالب بالا و نتایج به دست آمده مشاهده میکنیم که موسیقی بخش عظیمی از زندگی ما را تشکیل می دهد. و در اکثر فعالیت های ما تأثیر بسزایی دارد، و گوش دادن به آن باعث بهبود فعالیت های روزانه می شود. حال اگر بتوان این موسیقی را به نوعی مختص هر فعالیت با استفاده از هوش مصنوعی رایانه ای گسترش داد می توان تأثیر مثبت آن را بیشتر کرد. در نتیجه انسان ها هنگامی گوش دادن به موسیقی می توانند بازدهی بیشتر، درد کمتر، آرامش بیشتر و در کل حس ها را بهتر دریافت کرده و روز مثبتی داشته باشند. تحلیل هر موسیقی متناسب با مکان و زمان آن یکی از چالش هایی است که با توجه به بخش ۲-۲. می توان به آن پی برد. در بخش بعدی به مطالعه علم هوش مصنوعی رایانه ای می پردازیم تا بتوانیم این چالش را با استفاده از رایانه حل کنیم.

<sup>2</sup>Sympathetic nervous system

فصل سوم

هوش مصنوعی

### ۱-۳ نگاهی بر علم هوش مصنوعی

هوش مصنوعی هوشی است که توسط ماشین آلات در علوم کامپیوتر، زمینه تحقیقات هوش مصنوعی تعریف می شود. عنوان مطالعه «عوامل هوشمند»: هر وسیله ای که محیط خود را درک می کند و اقداماتی را انجام می دهد که شانس موفقیت آن را در یک هدف به حداکثر می رساند. اصطلاح عامیانه "هوش مصنوعی" زمانی اعمال می شود که یک ماشین کارکردهای "شناختی" که انسان با انسان دیگر مرتبط می کند تقلید کند. توانایی ها در حال حاضر به عنوان هوش مصنوعی طبقه بندی می شود که شامل درک موفقیت آمیز گفتار انسان است. ، رقابت در سطح بالایی در بازی استراتژیک، سیستم های (مانند شطرنج و گو)، ماشین های خودران و یاهوشمند مسیریابی در شبکه های تحویل محتوا، شبیه سازی های نظامی و تفسیر داده های پیچیده مشکلات (یا اهداف) باشند. اهداف بلند مدت تحقیقات هوش مصنوعی عبارتند از استدلال، دانش، برنامه ریزی، یادگیری، زبان طبیعی پردازش (ارتباط)، ادراک و توانایی حرکت و دستکاری اشیاء و هوش عمومی. هوش مصنوعی دارای دو روش برای تحلیل داده ها و یادگیری است که عبارتند از :

۱. یادگیر ماشین

۲. یادگیری عمیق

#### ۱-۱-۳ یادگیری ماشین

یادگیری ماشین زیرشاخه علوم کامپیوتر است که به گفته آرتور ساموئل در سال ۱۹۵۹، "کامپیوترها توانایی یادگیری بدون برنامه ریزی صریح را دارند." از مطالعه تشخیص الگو و نظریه یادگیری محاسباتی در هوش مصنوعی ، یادگیری ماشین مطالعه و ساخت الگوریتم هایی که می توانند از آنها یاد بگیرند و بسازند را ای را بررسی می کند. پیش بینی ها بر روی داده ها - چنین الگوریتم هایی بر موارد زیر غلبه می کنند دستورالعمل های برنامه کاملاً ثابت با ایجاد داده محور پیش بینی ها یا تصمیم گیری ها، از طریق ساختن یک مدل از ورودی های نمونه یادگیری ماشینی در طیف وسیعی از موارد استفاده می شود. وظایف محاسباتی که در آن طراحی و برنامه نویسی صریح است الگوریتم هایی با عملکرد خوب دشوار یا غیرقابل اجرا هستند. به عنوان مثال برنامه های کاربردی شامل فیلتر کردن ایمیل، تشخیص متجاوزان شبکه یا خودی های مخرب که روی یک داده کار می کنند. نقض، تشخیص کاراکتر نوری ، یادگیری به رتبه و بینایی کامپیوتری یادگیری ماشینی ارتباط نزدیکی ویا (اغلب همپوشانی دارد) با

آمار محاسباتی، که همچنین بر روی پیش بینی با استفاده از رایانه های قوی دارد ارتباط با بهینه سازی ریاضی، که روش ها را ارائه می دهد، حوزه های تئوری و کاربرد در این زمینه. فراگیری ماشین گاهی اوقات با داده کاوی ترکیب می شود، که بیشتر دومی است که در زیر زمینه بیشتر بر تجزیه و تحلیل داده های اکتشافی متمرکز است و می باشد به عنوان یادگیری بدون نظارت شناخته می شود. یادگیری ماشین نیز می تواند بدون نظارت باشد و برای یادگیری و ایجاد خط پایه استفاده شود نمایه های رفتاری برای موجودیت های مختلف و سپس برای یافتن آن ها استفاده می شود. ناهنجاری های معنی دار در حوزه تجزیه و تحلیل داده ها، یادگیری ماشینی یک روش مورد استفاده برای ابداع مدل ها و الگوریتم های پیچیده که خود را به پیش بینی وامیدارد. در استفاده تجاری، به عنوان تجزیه و تحلیل پیش بینی شناخته می شود. این مدل های تحلیلی اجازه می دهد محققان، دانشمندان داده، مهندسان و تحلیلگران برای «تولید تصمیمات و نتایج قابل اعتماد و قابل تکرار» و کشف «بینش پنهان». از طریق یادگیری از روابط تاریخی و روند در داده ها به آن ها دست یابند.

### ۲-۱-۳ یادگیری عمیق

یادگیری عمیق (همچنین به عنوان یادگیری ساختار یافته عمیق، یادگیری سلسله مراتبی یا یادگیری ماشین عمیق شناخته می شود). مطالعه ی شبکه های عصبی مصنوعی و یادگیری ماشینی مرتبط الگوریتم هایی که شامل بیش از یک لایه پنهان هستند است. نمونه هایی از شبکه های عمیق:

- از یک آبشار چند لایه غیربرای برای واحد های پردازشی استفاده میکند. تا بتواند خروجی لایه های قابل را به بعدی برساند. الگوریتم ما ممکن است نظارت شده، یا غیر نظارتی باشد.
- بر پایه الگوریتم غیر نظارتی چند مرحله برای نشان دادن داده است. اطلاعات مرحله های بالاتر از مرحله های پایین تر می رسد، تا ساختار سلسله مراتبی را کامل کند.

به طور خلاصه دو نوع از نورون های عصبی وجود دارد. نوع اول آن ها ورودی میگیرد، و نوع دیگری که خروجی می دهد. که بین این ها تعداد زیادی لایه های نورونی وجود دارند که خروجی های لایه قبل را پردازش کرده و به لایه بعد می دهند. تحقیقات در این زمینه تلاش میکنند تا نحوه نمایش داده های بزرگتر را بهتر کنند. بعضی از نمایش ها در این زمینه از علوم اعصاب و نحوه ارتباطی بین سلول های عصبی در مغز انسان الهام میگیرند. تکنیک های متفاوت یادگیری عمیق در زمینه های رایانه ای مانند تشخیص صدا، تشخیص سخنرانی، پردازش زبان طبیعی به کار گرفته می شوند.

## ۲-۳ خلاصه

هوش مصنوعی، یادگیری ماشینی و یادگیری عمیق اساساً ادراک ماشینی هستند. این قدرت تفسیر داده های حسگرها است. دو روش اصلی که ما چیزها را تفسیر می کنیم، این است که نامگذاری آنچه حس می کنیم. به عنوان مثال، صدایی می شنویم و با خودمان می گوییم "این صدای دخترم است." یا ما یک مِه از فوتون ها می بینیم و می گوییم "این صورت مادر من است." اگر این نام ها نباشند، هنوز هم می توانیم شباهت ها و تفاوت ها را تشخیص دهیم و شما ممکن است دو چهره را ببینید و بدانید که آنها مادر و دختر بودند، بدون اینکه نامشان را بدانید. یا شما ممکن است دو صدا را بشنوید و بدانید که آنها نوعی یکسان دارند. و لهجه آن ها یکسان است. الگوریتم ها برای نام گذاری از طریق یادگیری نظارتی آموزش می بینند. و خوشه بندی چیزها از طریق یادگیری بدون نظارتی آموزش می بینند. تفاوت میان یادگیری تحت نظارت و بدون نظارت این است که آیا شما یک آموزش با برچسب مجموعه ای برای کار با داده ها دارید یا خیر. برچسب هایی که اعمال می کنید به داده ها به سادگی نتایجی هستند که شما به آنها اهمیت می دهید. شاید شما به شناسایی افراد در تصاویر اهمیت می دهید. شاید برایتان مهم باشد که شناسایی ایمیل های خشمگین یا هرزنامه هایی که همه فقط حباب های بدون ساختار متن هستند. بنابراین یادگیری عمیق، کار با الگوریتم های دیگر طبقه بندی، خوشه بندی و پیش بینی است. این کار را با یادگیری سیگنال ها، یا ساختار، در داده ها به طور خودکار انجام می دهد. وقتی الگوریتم های یادگیری عمیق آموزش می بینند، آنها در مورد داده ها حدس می زنند، اندازه گیری خطای حدس های خود در برابر مجموعه آموزشی، و سپس راه حدس زدن را تصحیح می کنند تا به حدس های دقیق تر تبدیل شود. این بهینه سازی است. . حال تصور کنید که با یادگیری عمیق، می توانید طبقه بندی کنید، خوشه بندی کنید، یا پیش بینی کنید هر چیزی که راجع آن داده ای دارید، مانند: تصاویر، ویدیو، صدا، متن و سری های زمانی (بازارهای سهام، جداول اقتصادی و آب و هوا). یعنی هر چیزی که انسان ها می توانند حس کنند و فناوری ما می تواند آن را رایانه ای کند. شما توانایی تجزیه و تحلیل آنچه در دنیای اتفاق می افتد را با یادگیری عمیق چند برابر کرده اید. ما اساساً به جامعه توانایی این را می دهیم تا بسیار بیشتر هوشمندانه باشد. پیش بینی به تنهایی قدرت عظیمی است، طبقه بندی پیش پا افتاده به نظر می رسد، اما با نامگذاری چیزی، شما می توانید تصمیم بگیرید که چگونه به آن پاسخ دهید. اگر یک ایمیل است، هرزنامه، آن را به پوشه هرز ها ارسال می کنید و در زمان خواننده صرفه جویی می کنید. اگر چهره ای که توسط دوربین درب ورودی شما ثبت شده مادر شماست، شاید به قفل هوشمند بگویید در را باز کند. اگر اشعه ایکس یک الگوی تومور را نشان می دهد، شما آن را برای بررسی عمیق

تر علامت گذاری می کنید تا آن را درمان کنید. در نتیجه کارهایتان را می توانید با هوش مصنوعی آسان تر کنید. [۱۳]

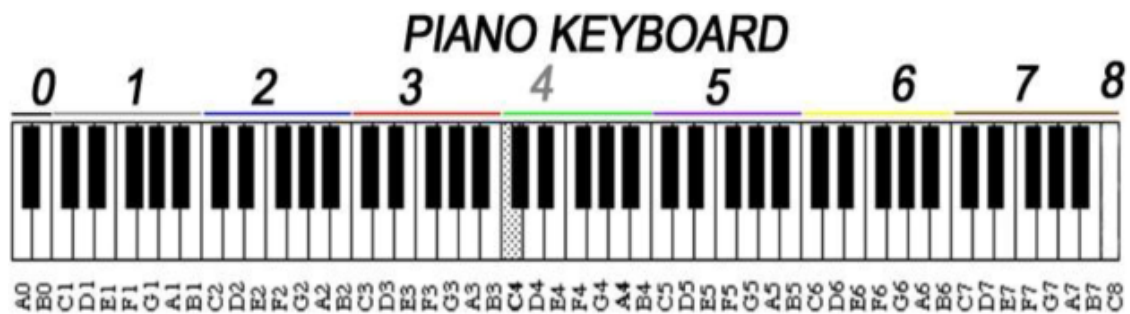
## فصل چهارم

### نحوه تبدیل موسیقی به زبان رایانه



## ۱-۴ مقاطعات پس زمینه موسیقی

درک اینکه چگونه مولد موسیقی ما کار می کند نیاز به دانش پیش زمینه در برخی از مفاهیم اساسی دارد. تئوری موسیقی. این بخش شامل خلاصه ای بسیار کوتاه از اصطلاحات ضروری موسیقی است. اصطلاح گام صدا برای توصیف بلندی یا پایین بودن صدا استفاده می شود. در موسیقی هفت نت (یا "صدا") وجود دارد که A, B, C, D, E, F, G. هر مجموعه نت (A تا G) متعلق به یک هشتایی<sup>۱</sup> هستند. از هشتایی برای تعیین این استفاده می شود که تعیین کنیم که نت ها در کدام گام اند. برای مثال، A1 (A در هشتایی ۱) بسیار عمیق تر و پایین تر از A7 (A در ۷) خواهد بود. هشتایی ۷ که گام بالاتر و صدای تندتر است. یک کیبورد معمولی پیانو شامل ۸۸ کلید قابل پخش و هفت هشتایی کامل همانطور که با رنگ های شکل ۱-۴ نشان داده شده است. اکثر قطعات موسیقی در یک کلید هستند.



شکل ۱-۴ کیبورد پیانو

یک کلید، یادداشتی است که پایه (یا "خانه") را تشکیل می دهد A موسیقی در قطعه یک قطعه نواخته شده در کلید E به این معنی است که نت E موسیقی در آهنگ است. ترازو مجموعه خاصی از یادداشت ها است که بر پایه کلید ساخته شده اند. این را می توان به عنوان یک مجموعه ابزار خاص توصیف کرد یادداشت هایی که با آن می توان آهنگ ایجاد کرد. ترازوها با یکدیگر متفاوت هستند از این نظر که برخی برای موسیقی "شادتر" یا "روشن تر" (در مقیاس C ماژور) و برخی برای "غمگین" بهتر هستند. یا قطعات "تاریک" مانند مقیاس های کوچک D یا E. آکورد گروهی از نت ها است که به طور همزمان نواخته می شوند و صدایی متمایز را به صورت گروهی تولید می کنند. مثلا، آکورد Major A از نت های A, C, و E تشکیل شده است. در موسیقی، جابجایی فرآیند "تغییر یک ملودی است، یک پیشرفت هارمونیک یا یک قطعه موسیقی کامل به یک کلید دیگر" در حالی که همان ساختار لحن را حفظ می کند». این به این معنی است که ساختار آهنگ در حین تغییر کلید حفظ می شود.

<sup>1</sup>Octave

## ۴-۱-۱ نوع داده

فرمت رابط دیجیتال ابزار موسیقی میدی ( MIDI یا .mid ) برای ذخیره دستورالعمل های پیام استفاده می شود. شامل نت، فشار/حجم، سرعت، زمان شروع و پایان آنها، عبارات و غیره است. صدا را مانند فرمت های صوتی ذخیره نمی کند، بلکه دستورالعمل هایی را در مورد نحوه تولید صدا ذخیره می کند. این دستورالعمل ها را می توان توسط یک کارت صوتی که از یک جدول موج (جدول صدای ضبط شده) استفاده می کند، تفسیر کرد. امواج) برای ترجمه پیام های MIDI به صدای واقعی. این ها توسط نرم افزار استودیو پخش کننده midi مانند Studio (FL) Loops Fruity یا ترتیب دهنده های محبوب مانند سنتزیا ( Synthesia ) (نگاه کنید به شکل ۴-۲) نیز قابل تفسیر هستند. نرم افزارهای نت نویسی موسیقی مانند موس اسکور ( MuseScore ) یا فاینال ( Finale ) می توانند midi را به نت های قابل ویرایش ترجمه کنند. این به کاربران این امکان را می دهد که موسیقی را با نماد موسیقی معمولی روی رایانه خود بنویسند و سپس با MIDI به آن گوش دهند مانند یک نوازنده بدون نیاز به ساز. پیام



شکل ۴-۲ نمونه یک تصویر از صفحه برنامه

های MIDI به پیام های سیستم و کانال جدا می شوند. از آنجایی که ما در درجه اول نگران هستیم در یادداشت ها، ما فقط روی پیام های کانال تمرکز می کنیم. ۱۶ کانال MIDI ممکن (۱۶ جریان از پیام های کانال حاوی یادداشت هایی برای پخش). هر پیام کانال MIDI حاوی نوع پیام است. پیام، زمان و شماره یادداشت با این حال، پیام ها ممکن است حاوی اطلاعات دیگری نیز باشند این مقاله تنها پیام های یادداشت اولیه در نظر گرفته شده است. بخش عمده ای از پیام های MIDI عمدتاً یادداشت ها و خاموش ها هستند. یک یادداشت نشان می دهد که یک کلید در حال فشار دادن است (یک نت در حال پخش است). آن یادداشت ادامه خواهد داشت برای پخش تا زمانی که پیام یادداشت آف دریافت شود، که نشان می دهد کلیدی دیگر فشار داده نمی شود. در داخل پیام کانال شماره یادداشت (یعنی  $C = 60$  وسط) و داده های دیگر مانند سرعت (حجم توجه داشته باشید). شایان ذکر است که برخی از پیام های MIDI با نوع داده یادداشت و سرعت  $0 =$  به عنوان یک نوشته می شوند. جایگزینی برای نت آف، با این

حال هر دو به این معنی است که پخش نت متوقف شده است. در شکل ۴-۳، اولین پیام نشان می دهد که نتی با گام = ۵۰ و سرعت = ۵۲ در حال پخش است. زمان در پیام‌های یادداشتی صفر است زیرا زمان نشان می‌دهد که کلید چقدر زمان پخش شده است. را پیام یادداشت پس از آن نشان می دهد که یادداشت ۷۲ پس از گرفتن زمان ۰.۶۷ متوقف شده است (زمان midi سفارشی، نه ثانیه). پس از آن، یک یادداشت روی پیام، note = ۷۱ را پخش می‌کند. سپس توسط پیام یادداشت آف متوقف می‌شود. مدت زمانی که این یادداشت طول کشید ۰.۲۲۳ بود.

```
note_on channel=5 note=50 velocity=52 time=0
note_off channel=1 note=72 velocity=0 time=0.6716414999999999
note_on channel=1 note=71 velocity=55 time=0
note_off channel=1 note=71 velocity=0 time=0.22388049999999998
```

شکل ۴-۳ نمونه یک پیام

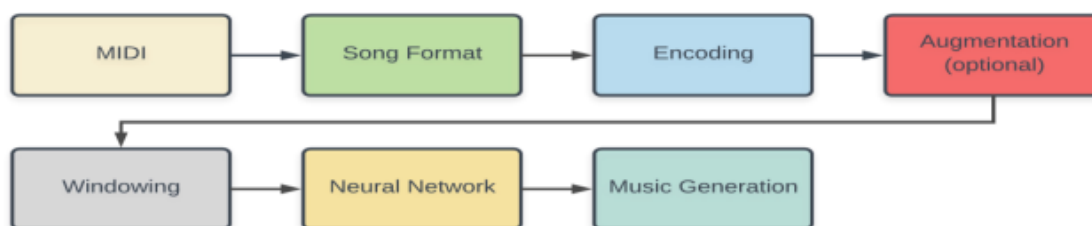
## ۴-۱-۲ شبکه‌های عصبی

یک شبکه‌های عصبی به عنوان "مجموعه به هم پیوسته عناصر پردازش ساده (واحدها یا گره‌ها)" تعریف می شود. پردازش در شبکه به وزن اتصالات بین گره‌هایی که توسط آنها آموزش داده می شود متکی است. یا با مجموعه داده آموزشی تطبیق داده شود. این فرآیند همان چیزی است که اغلب آموزش شبکه یا یادگیری نامیده می شود. شبکه‌های عصبی ها ابزاری ایده آل برای مسائل طبقه بندی هستند. شبکه‌های عصبی ها تطبیقی هستند زیرا با توجه به داده های ورودی تنظیم می شوند. آنها قادر هستند شناسایی توابعی که ویژگی ها را با درجه ای از دقت به دسته ها مرتبط می کنند. آن ها به اندازه کافی انعطاف پذیر هستند تا مشکلات پیچیده طبقه بندی دنیای واقعی مانند پیش بینی ورشکستگی، تصویر و تشخیص گفتار و غیره را تشخیص دهند. با این حال، همه شبکه‌های عصبی ها یکسان نیستند. شبکه‌های عصبی ها به دو دسته کلی تقسیم می شوند: پیشخور، و بازگشتی. در پیشخورها شبکه‌های عصبی فعال سازی از طریق لایه ورودی به لایه خروجی بدون هیچ گونه بازگشتی جریان می یابد. فقط یک لایه پنهان در آن وجود دارد بین آنها نام شبکه‌های عصبی پیشخور "عمیق" به این معنی است که لایه های پنهان بیشتری وجود دارد، اما با همان مفهوم. ماهیت شبکه‌های عصبی پیشخور برای تولید موسیقی نامناسب است زیرا حافظه ندارد. هیچ حالت گذشته شبکه‌های عصبی نمی تواند روی حالت های آینده وزن ها در شبکه تأثیر بگذارد. نواهای های موسیقی اساسا متنی هستند و برای ایجاد نت های جدید به حافظه نت های گذشته متکی هستند. منسجم و آهنگین (صدای دلپذیر).

به همین دلیل است که شبکه‌های عصبی پیشخور برای تولید کننده های موسیقی کاملاً نامناسب هستند.

## ۲-۴ معماری پیشنهادی

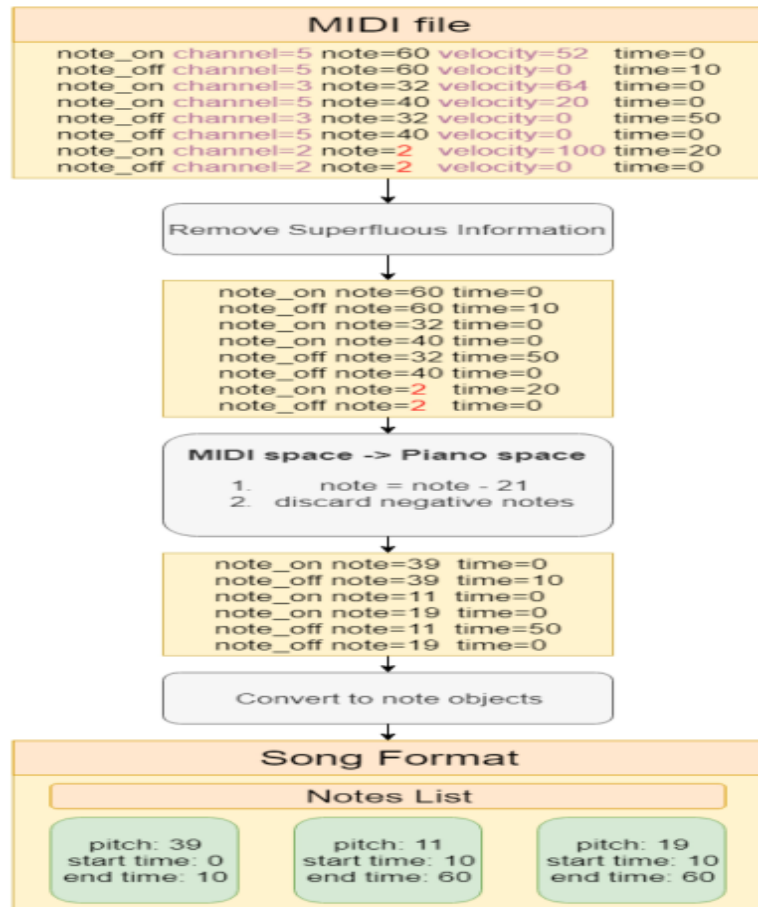
در این مقاله، جریان نشان داده شده در شکل ۴-۴ را پیشنهاد می کنیم.



شکل ۴-۴ نقشه راه پیشنهادی

## ۱-۲-۴ تبدیل داده ها به نوع آهنگ

تمامی فایل های midi به فرمت واسطه ای (فرمت آهنگ) که ما توسعه داده ایم، تبدیل شده اند به ما کمک می کند تا با نت ها به روشی طبیعی تر برخورد کنیم. در حالی که روی رویدادهای هر فایل حلقه می زنیم، یادداشت ها هستند پخش شده در تمام ۱۶ کانال MIDI در قالب آهنگ ضبط می شود. شایان ذکر است که گاهی اوقات فایل های MIDI از پیام «note\_on» با «سرعت» ۰ برای نشان دادن «note\_off» استفاده می کنند. هر رویداد MIDI حاوی یک ویژگی زمان ("time") است که عددی مربوط به زمان (بر حسب تیک) از آخرین پیام را در خود دارد. فرمت آهنگ مجموعه ای از دو کلاس است که وظیفه نگهداری اطلاعات نت و تبدیل به آن را دارند. و از فایل های MIDI. بنابراین، قابلیت صادرات و وارد کردن MIDI به فرمت کدگذاری شده انجام می شود. داخل کلاس فرمت آهنگ (Format Song) . علاوه بر این، از آنجایی که MIDI دارای محدوده نت [۰، ۱۲۷] است، و ما فقط علاقه مند هستیم در زیر مجموعه ای که کلیدهای پیانو هستند (زیر مجموعه [۲۱، ۸۸ + ۲۱] در MIDI)، فرآیند تبدیل فرمت آهنگ این تغییر از فضای MIDI به فضای پیانو را انجام می دهد. این فرآیند در شکل ۴-۵ نشان داده شده است.



شکل ۴-۵ فرایند تبدیل نوع آهنگ

## ۴-۲-۲ رمزگذاری

از آنجایی که هیچ راهی برای تغذیه شبکه با مقادیر MIDI به طور مستقیم و توانایی تولید موسیقی وجود ندارد، رمزگذاری داده های موسیقی در قالبی که توسط شبکه قابل استفاده باشد بسیار مهم است. خروجی از روش رمزگذاری مورد استفاده، ماتریسی از بردارهای باینری با اندازه  $N \times 89$  برای هر فایل MIDI است که  $N$  یک عدد است. بردارها در قطعه ۸۸ بیت اول در بردار با وضعیت ۸۸ نت صفحه کلید پیانو مطابقت دارد. صفر نشان می دهد که یک نت در این لحظه پخش نمی شود، در حالی که یک نشان دهنده عکس آن است ۸۹ بیت در بردار تعداد تکرار است. تعداد تکرار به ما می گوید که این بردار چند بار تکرار شده است. یعنی چه مدت این وضعیت فعلی صفحه کلید حفظ می شود تا زمانی که تغییر کند. در آزمایشات اولیه ما، تعداد تکرارها یک عدد طبیعی بود. با این حال، داشتن یک عدد طبیعی که نرمال نشده باشد، محاسبه ضرر و همگرایی را به هم می ریزد. بنابراین، در پایان از یک عدد واقعی در محدوده  $[0, 1)$  استفاده کردیم. جایی که عدد نزدیک به ۰ به معنای تکرار کمتر است و ۱ به معنای بیشترین

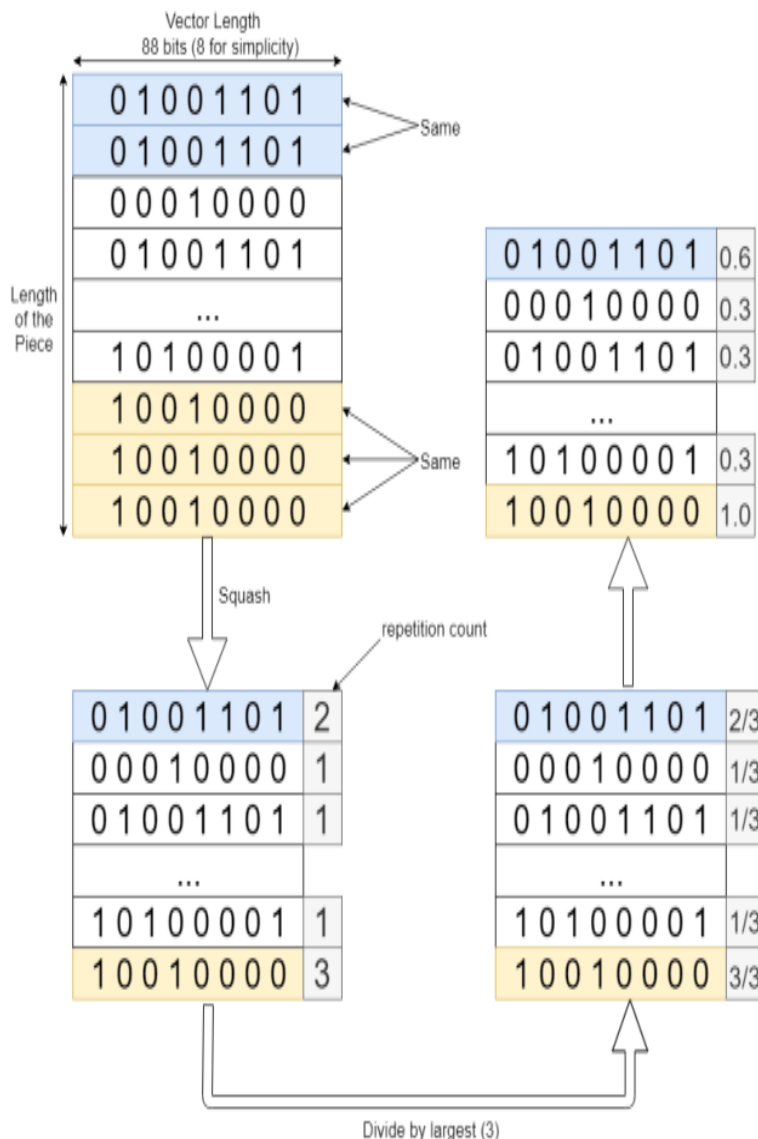
مقدار در داده MIDI است. فواصل نیمه باز هستند زیرا هیچ برداری نباید مقدار تکرار صفر داشته باشد، زیرا از نظر معنایی اشتباه است. ما آگاهانه تصمیم گرفتیم که تعداد تکرار را به جای عمومی شدن، محلی کنیم. یعنی ۱ بزرگترین باشد. توجه داشته باشید در فایل و نه در کل مجموعه داده. استدلال ما این است که تعداد تکرار از نظر معنایی پاسخ می دهد به سوال «این نت نسبت به سایر نت های موجود در فایلش چه مدت پخش می شود؟» و این عدد نباید توسط بردارهای دیگر در قطعات دیگر تحت تأثیر قرار گیرد. ، حتی اگر همه این بردارها وارد فرآیند آموزش شوند. یعنی نت بلند در فایل نت های کوتاه نباید ارزشی کمتر از نت بلند در فایل نت های بلند داشته باشد. شکل ۴-۶ روش رمزگذاری مورد استفاده را نشان می دهد و نشان می دهد که بردارهای تکراری (بردارهای رنگی) چگونه به دست می آیند. با هم ترکیب شده عدد در تعداد تکرار (خاکستری) ثبت می شود. تعداد تکرارها با تقسیم آن بر بزرگترین تعداد تکرار عادی می شود.

### ۴-۲-۳ افزایش داده ها

یک نوای خاص را می توان با هر کلید متفاوت با ندا های بسیار متفاوت نواخت اما معنایی یکسان دارد. معنی و به عنوان همان نوا شنیده شود. شکل ۴-۷ در ندا موسیقی نشان می دهد که چگونه همان نوا می تواند . با دو کلید موسیقی نواخته شود. در خط اول، نوای با کلید موسیقی "سی ماژور" پخش می شود، در حالی که در خط دوم اینکه با کلید موسیقی "ف ماژور" پخش می شود. این دو ملودی اگر به طور جداگانه شنیده شوند، معادل هستند اگرچه آنها کاملاً نت های متفاوتی دارند. شکل ۴-۹۷ برای حل این مشکل از تقویت کننده روی قطعات ما استفاده شده است.

همانطور که در شکل ۴-۸ می بینیم، پس از رمزگذاری، از هر نوا کپی هایی می سازیم و آنها را به طور تصادفی با انجام یک شیفت به سمت راست جابجا می کنیم. یعنی کلید هر همانند را طوری تغییر می دهیم که تعداد مشخصی همانند بدست آید. همانند های متمایز هر کدام در یک کلید متفاوت. دلیل استفاده از جایگشت های تصادفی این است که بتوانیم کارهای کمتری انجام دهیم. بیش از ۱۱ افزایش بدون تغییر دادن کردن کلیدهای مجموعه داده. یعنی اگر سه تا افزایش بدون جهش اولیه انجام دادیم با جابجایی، در نهایت به داده هایی می رسیم که به کلیدهای ۱+، ۲+ و ۳+ منتقل می شوند و این باعث ایجاد یک عدم تعادل در مجموعه داده که در آن کلیدهای [۴+، ۱۱+] کمتر نمایش داده می شوند.

<sup>2</sup>Long-short term memory in RNN



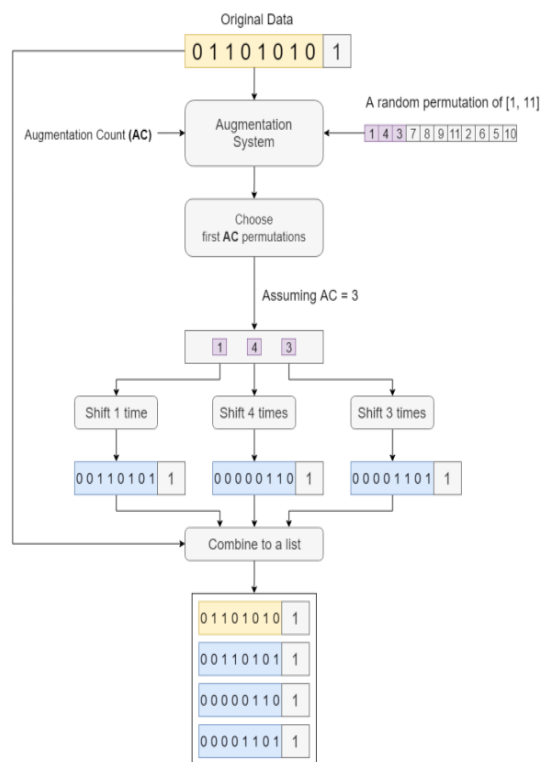
شکل ۴-۶ روش کد گذاری



شکل ۴-۷ گذر از نوت سی ماژور به اف ماژور

## ۴-۲-۴ گام های زمانی

از آنجایی که ل.س.ت.م<sup>۲</sup> ها انتظار دارند که شکل ورودی در قالب (گام های زمانی، تعداد ویژگی ها) باشد، باید ما داده ها را تغییر دهیم. داده ها باید در این فرمت باشند. برای انجام این کار از روش "پنجره



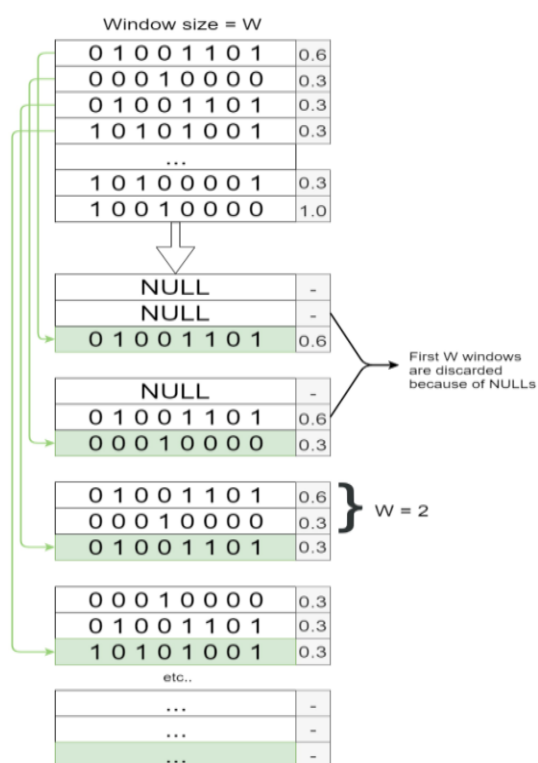
شکل ۴-۸ نمونه یک مقاله در گوگل اسکولار

کشویی” استفاده می کنیم تا در هر نمونه در آموزش، نمونه شامل خود و  $N$  نمونه قبلی است که در شکل ۴-۹ نشان داده شده است. [۱۴]

## ۳-۴ خلاصه

به طور خلاصه در ۴-۱ این بخش نحوه خواندن موسیقی و تبدیل آن به زبان رایانه را دیدیم. که چگونه نواها و صداهای یم اهنگ موسیقی را به زبان رایانه تبدیل کنیم. چگونه مشکلات ایجاد شده در حین تبدیل کردن را بر طرف کنیم. و چگونه داده های به دست آورده را توسط یک شبکه عصبی تحلیل کرده و به ساخت موسیقی پردازیم. در بخش بعدی نیز نحوه تبدیل موسیقی هایی که به زبان رایانه تبدیل شدند را به موسیقی جدید و نحوه ساخت آن را خواهیم دید.





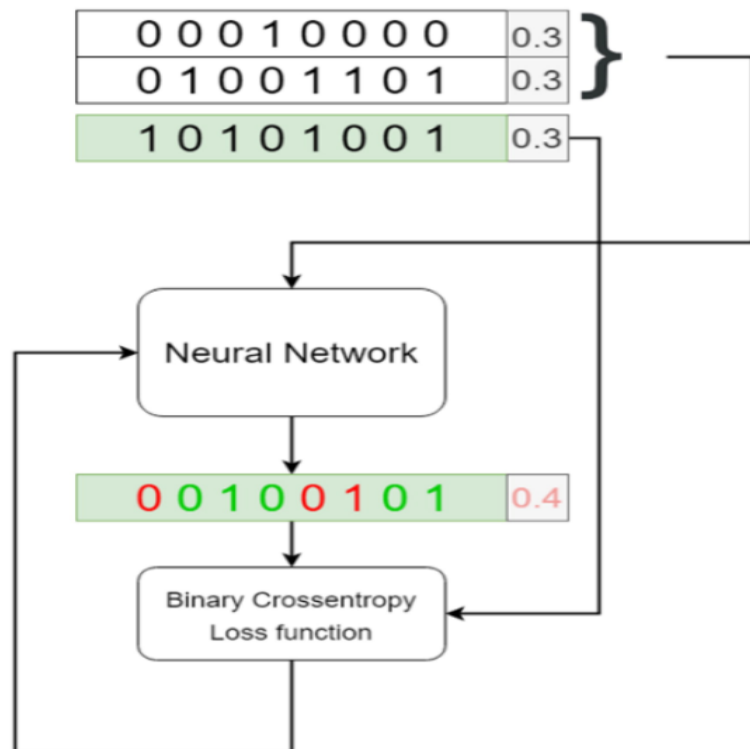
شکل ۴-۹ نمونه یک مقاله در گوگل اسکولار

## فصل پنجم

### نحوه ساخت موسیقی

## ۱-۵ شبکه عصبی

با داشتن مجموعه داده رمزگذاری شده پنجره ای، داده ها به شبکه عصبی وارد می شود. نمونه ها و برچسب ها را طوری جدا می کنیم که در نمونه  $n$  مثالی با گام های زمانی آن در محدوده  $[n-w, n-1]$  می آوریم و اجازه می دهیم شبکه عصبی . نمونه را در  $n+1$  پیش بینی کند. شکل ۱-۵ از پنجره ای از داده های تولید شده از پنجره سازی در شکل ۴-۹ استفاده می کند و نشان می دهد که چگونه این پنجره به شبکه تغذیه می شود.

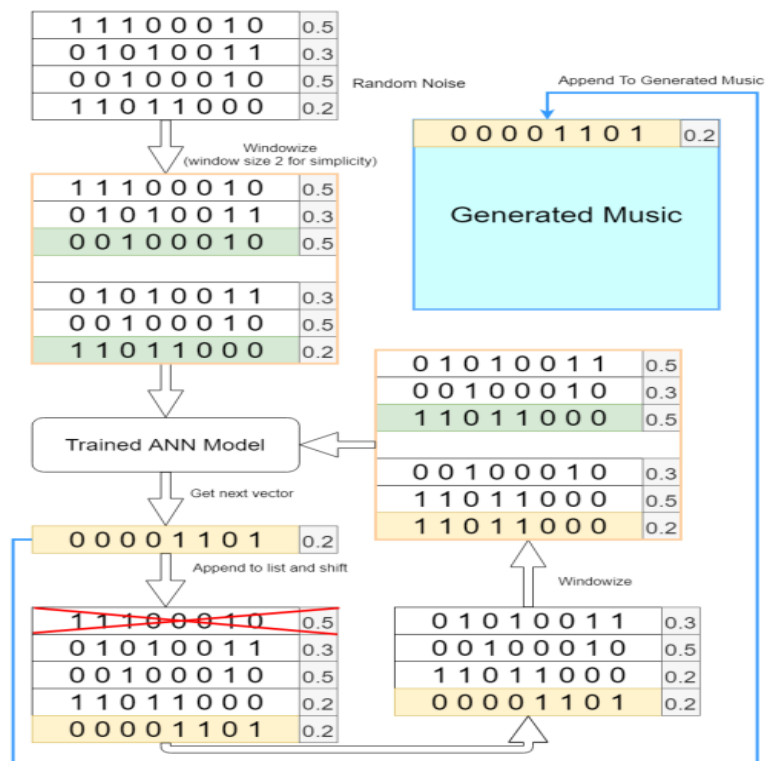


شکل ۱-۵ آموزش شبکه عصبی

## ۲-۵ تولید موسیقی

در فواصل زمانی ثابت در طول آموزش شبکه، شبکه موسیقی تولید می کند و سپس آن را با نوع MIDI ذخیره می کند. به منظور گوش دادن به آن و تجزیه و تحلیل آن. شکل ۲-۵ نشان می دهد که چگونه موسیقی پس از آموزش آن تولید می شود. برای تولید موسیقی، شبکه با یک ماتریس اختلال یا یک ماتریس تصادفی از مجموعه داده و از آن بخواهید بردار بعدی را پیش بینی کند. پس از پیش بردار بعدی، این بردار را به ماتریس اضافه می کنیم، و همچنین لیست بردارهای تولید شده و آخرین بردار را از

ماتریس خارج می‌کنیم. به همین ترتیب، همان فرآیند برای پیش بینی بردار بعدی انجام می‌شود. این کار چندین بار تکرار می‌شود تا یک تکه از طول مورد نظر به دست می‌آید.



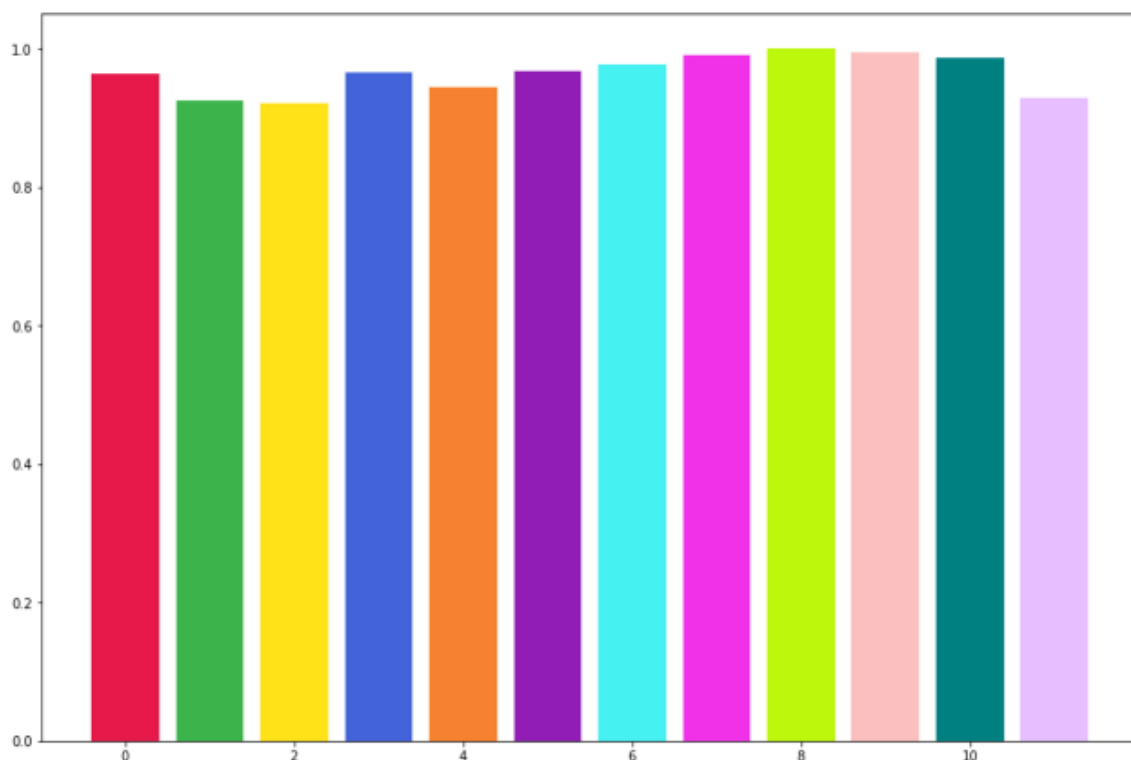
شکل ۵-۲ روش تولید موسیقی

## ۳-۵ داده ها و ابزار ها

### ۱-۳-۵ مجموعه داده ها

« کتاب کلاویه خوش خلق و خوی II » باخ به عنوان مجموعه داده ما انتخاب شده است. این به دلیل سبک سیستماتیک موسیقی باخ و همچنین ماهیت خاص این اثر است که باخ در آن برای هر کلید دو قطعه موسیقی می‌نویسد. مجموعه داده شامل ۲۴ فایل است که هر فایل شامل دو قطعه است. شکل ۳-۵ یک هیستوگرام نرمال شده از تعداد ۱۲ کلید در مجموعه داده را نشان می‌دهد. هر وقوع یک کلید بدون توجه به هشتایی شمرده می‌شود. این نشان می‌دهد که همه کلیدها با همان درجه رخ می‌دهند، یعنی هیچ کلید غالب در مجموعه داده وجود ندارد. انحراف استاندارد نمودار میله‌ای عادی شده ۰.۲۶۸۳۰ است. این نشان می‌دهد که به طور متوسط تفاوت در تعداد وقوع بین هر دو کلید در مجموعه داده حدود ۶.۲ درصد است. شکل ۵-۴ وقوع هر یک از ۸۸ یادداشت را نشان می‌دهد. ما می

توانیم مشاهده کنیم که بیشتر نت ها در محدوده میانی پیانو قرار دارند که معمولاً در همه موسیقی ها مورد انتظار است. با توجه به این ویژگی ها، این مجموعه داده به خوبی متعادل است و کاملاً با نیازهای ما مطابقت دارد.



شکل ۵-۳ شماره کلید های مجموعه داده ها محور افقی شماره کلید و محور عمودی فرکانس عادی شده

## ۵-۳-۲ ابزار های استفاده شده

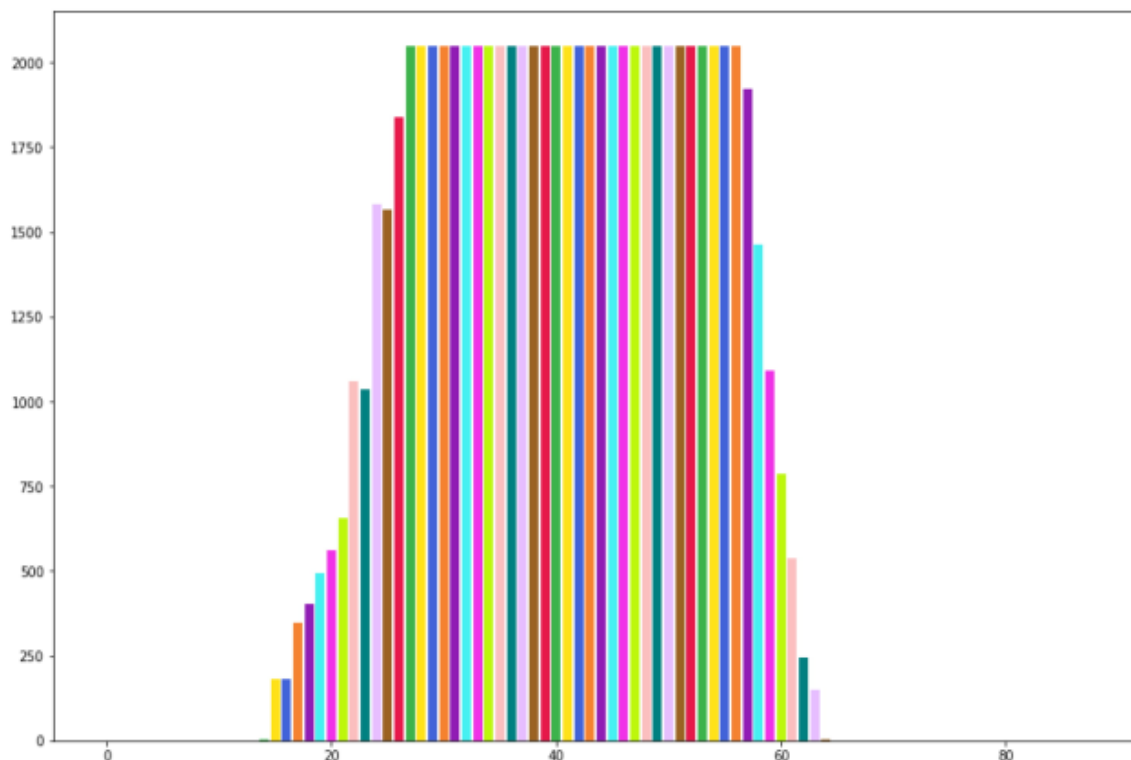
ابزارهای مورد استفاده برای اجرای رویکرد پیشنهادی عبارتند از:

- کراس<sup>۱</sup> در باطن تنسورفلو<sup>۲</sup>. کراس یک کتابخانه سطح بالا برای یادگیری ماشین با شبکه عصبی است که به ما اجازه می دهد تا به سرعت نمونه سازی کنیم و پارامترها را بدون بازنویسی تعداد زیادی کد تغییر دهیم. تنسورفلو گوگل که کتابخانه ای برای محاسبات روی نمودارها است که کاملاً با شبکه های عصبی مطابقت دارد و به طور موثر بر روی کارت گرافیک اجرا می شود.
- تخته تنسور<sup>۳</sup>. ابزاری از تنسورفلو برای یادگیری تجسمی است.

<sup>۱</sup>keras

<sup>۲</sup>Tensorflow

<sup>۳</sup>TensorBoard



شکل ۴-۵ محور افقی شماره کلید پیانو و محور عمودی فرکانش عادی نشده

- نام پای ۴. نام پای یک کتابخانه علمی جبری خطی است که به طور موثر با ماتریس ها و بردارها سروکار دارد این کتابخانه در کراس استفاده می شود و داده های آموزشی و آزمایشی باید آرایه های نام پای قبل از تغذیه آنها به هر شبکه باشد. علاوه بر این، ما از نام پای در بخش رمزگذاری و رمزگشایی برای مقابله با آن استفاده می کنیم. ماتریس ها (به عنوان مثال افزودن یک بردار به چند ردیف در یک ماتریس).

- میدو ۵. میدو یک کتابخانه پایتون است که به خواندن فایل های MIDI و نوشتن آنها کمک می کند.

- دفتر ژوپیتر ۶. یک برنامه وب برای تجسم داده ها و مدل سازی آماری. ما از دفترچه ژوپیتر استفاده می کنیم تا در آزمایش بتوانیم بلوک های کد را در هر دنباله ای که انتخاب می کنیم اجرا کنیم. این به نمونه سازی سریع آزمایش ها و مدل های شبکه عصبی کمک می کند.

<sup>4</sup>NumPy

<sup>5</sup>Mido

<sup>6</sup>Jupyter NoteBook

### ۵-۳-۳ شبکه عصبی

شبکه عصبی در کراس ایجاد شده است و آن را بر روی داده های به دست آمده از فرآیند پنجره سازی آموزش می دهد. همانطور که در شکل ۵-۵ نشان داده شده است، شبکه عصبی از یک یا چند لایه ال.اس.تی.ام (LSTM) تشکیل شده است که هر یک از آنها یک لایه انصرافی به دنبال دارد که کمک می کند. تمرین را منظم کنید و از اضافه کردن آن جلوگیری کنید. روال انصراف به این صورت است که برخی از گره ها را به صورت تصادفی رها می کند. به دنبال آن یک عملیات مسطح انجام می شود که ابعاد ماتریس را از سه بعدی به عنوان خروجی کاهش می دهد. لایه های LSTM را به دو بعدی برای لایه های کاملاً متصل تبدیل می کند. سپس لایه های کاملاً متصل را اضافه می کنیم و هر کدام را به دنبال آن اضافه می کنیم. یک لایه ترک تحصیل در نهایت، خروجی شبکه عصبی یک لایه کاملاً متصل با اندازه ۸۹ با یک سیگموئید<sup>۷</sup> است. تابع فعال سازی روی آن اعمال می شود. برای راهنمایی آموزش شبکه عصبی، ما از بی نظمی<sup>۸</sup> متقاطع دودویی استفاده کردیم. تابع ضرر (Log) (Loss) به جای میانگین مربعات خطای کلاسیک. این به دلیل ماهیت ماست که نزدیک به یک مسئله طبقه بندی چند برچسبی با ۸۸ برچسب و در چند برچسبی است که طبقه بندی، دودویی متقاطع بی نظمی امتیاز ضرر را بیشتر از دقت مدل معمولی می دهد. به دلیل اینکه بیت ۸۹ یک مقدار واقعی است و یک مقدار دودویی نیست، این تابع از دست دادن کامل نیست با این حال، برای هدف ما به اندازه کافی نزدیک است.

برای قضاوت در مورد عملکرد شبکه، داده ها را به ۷۰ درصد داده های آموزشی و ۳۰ درصد اعتبار سنجی تقسیم می کنیم. سپس از اندازه گیری اف<sup>۹</sup> که میانگین هارمونیک دقت و یادآوری است برای محاسبه دقت شبکه استفاده می کنیم.

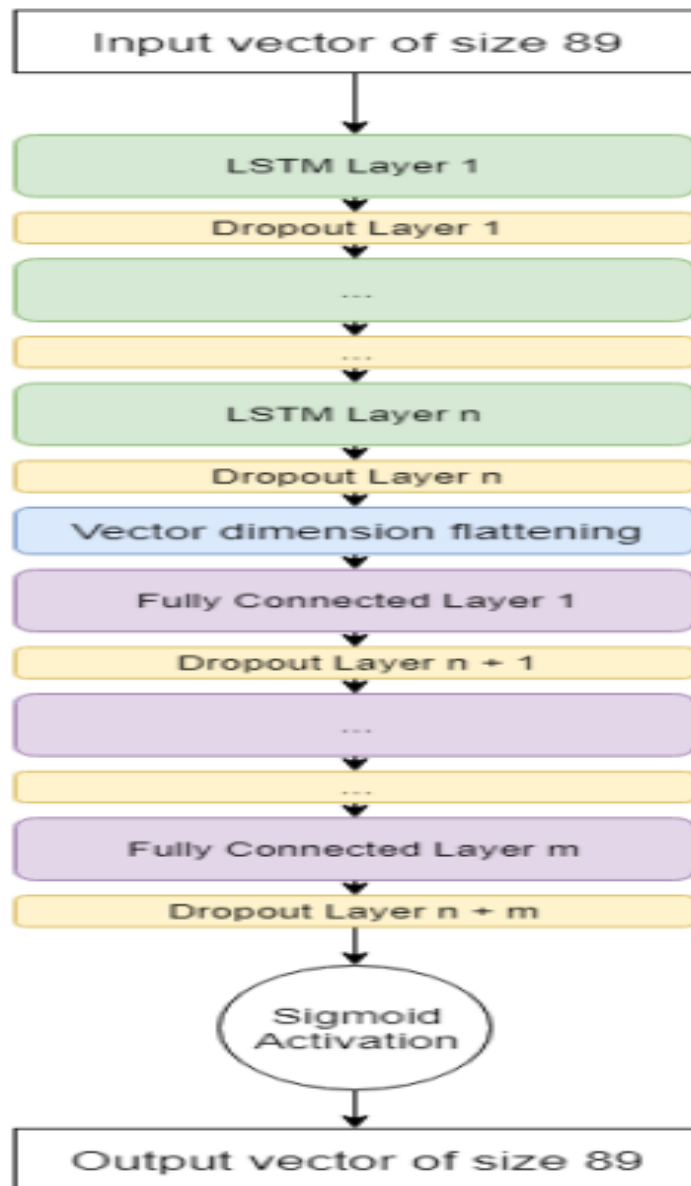
### ۵-۴ نتایج

در این فصل نحوه ساخت موسیقی با استفاده از موسیقی که در فصل گذشته به زبان رایانه تبدیل کرده بودیم را دیدیم. که با استفاده از ابزار های متفاوت چند موسیقی را برای آموزش به یک شبکه عصبی می دهیم و موسیقی های جدیدی را تولید می کنیم. و با تابع های متفاوت درستی آن را می سنجیم. اما در اینجا چالش اصلی ما این است که موسیقی به دست آمده برای مخاطب نیز تاثیر گذار باشد. و آیا

<sup>7</sup>Sigmoid activation function

<sup>8</sup>Entropy

<sup>9</sup>F-measure



شکل ۵-۵ معماری شبکه عصبی

شبکه عصبی استفاده شده بهترین حالت است یا خیر؟ در فصل بعد نتایج به دست آمده از مشاهدات خود را میبینیم و راهی برای از بین بردن این چالش ها ارائه میدهم. [۱۴]



## فصل ششم

### نتیجه‌گیری و پیشنهادات

## ۱-۶ نتیجه‌گیری

در این مطالعه، ما سعی کردیم با توسعه یک رویکرد، رفتار تولید موسیقی را بررسی کنیم. متشکل از شش مرحله، یعنی تبدیل فرمت به فرمت آهنگ، رمزگذاری، تقویت، پنجره‌سازی، یادگیری با استفاده از شبکه عصبی و تولید موسیقی. آزمایش‌های مختلفی با پارامترهای مختلف برای بررسی بهبود یادگیری نتایج ایجاد شده توسط شبکه ما بود از منظر موسیقی خیلی چشمگیر نیست، اما بد هم نیست. مدل ما یک درک اساسی از ریتم و هارمونی قطعات تولید شده ساختار بزرگی ندارند. آغاز، پایان و موتیف تکرار برخی از این موارد می‌توانست توسط شبکه‌ای شبیه به مدل والد مورد توجه قرار گیرد. برخی از مشکلاتی که حل نشده باقی می‌مانند این است که مدل در موارد خاص چند نت خوب را خروجی می‌دهد. سپس تمام صفرها همچنین گاهی اوقات ملودی‌های کوتاهی تولید می‌کند که بی‌نهایت تکرار می‌شوند. این دو مشکل نیاز دارند به طور عمیق مورد بررسی قرار گیرد و ممکن است یک فرآیند پست برای جلوگیری از وقوع چنین مواردی به ندرت اجرا شود موارد اتفاق افتاده مشکل سوم این بود که تابع از دست دادن بیت ۸۹ بودن را در نظر نمی‌گیرد ارزش شناور اگر دو مشکل اول را حذف کنیم زیرا همیشه رخ نمی‌دهند، عملکرد مدل از نظر موسیقی قابل قبول است و با صدای کمتر هارمونی خوبی می‌دهد. برای مشکل سوم، ساختار باید با اتصال مستقیم ورودی به خروجی به بیت ۸۹ اهمیت بیشتری بدهد. می‌تواند تأثیر واضحی بر شبکه عصبی و در تابع ضرر داشته باشد. کار احتمالی آینده برای کار ما می‌تواند استفاده از توابع از دست دادن سفارشی باشد که پراکندگی را بهتر مدل می‌کند. ماهیت بردارها همچنین می‌توانیم از یک بیت ۹۰ اضافی برای نمایش موقعیت درون قطعه موسیقی استفاده کنیم. که می‌تواند به معرفی ساختار بزرگتر در قطعات تولید شده کمک کند. یکی دیگر از بهبودهای احتمالی استفاده از مدل‌هایی غیر از این مانند شبکه‌های متخاصم مولد. ما همچنین می‌توانیم با انواع دیگر مجموعه داده‌های موسیقی با ویژگی‌های متفاوت با مجموعه‌ای که ما استفاده کردیم آزمایش کنیم. بالاخره مشکل می‌تواند به فایل‌های چند ابزاری و تولید چند ابزاری گسترش یابد. امیدوارم این‌ها امکان پذیر باشد پیشرفت‌ها می‌تواند الهام‌بخش آزمایش‌های بیشتر توسط دیگران باشد. هر چند که شنیدن موسیقی زنده لذت بسیار زیادی را دارد و تا آینده‌ای دور نمی‌توان جای آن را با موسیقی تولید شده با رایانه گرفت. اما شاید در آینده‌ای دور این روش جایگزین انسان‌ها شود این روش بسیار دلنشین‌تر برای انسان‌های نسل جدید باشد.

## ۲-۶ پیشنهادات

با استفاده از موسیقی‌های تولید شده توسط هوش مصنوعی می‌توان موسیقی را در صنایع مختلف گسترش داد. برای مثال اگر رفتار مشتریان در یک فروشگاه را تحلیل کرد و با استفاده از آن‌ها موسیقی‌های جدید تولید کرد که به سمت هدف خاصی در خرید مشتریان را هدایت کرد. در موردی دیگر در فیلم‌ها می‌توان با توجه به اینکه احساسات بیننده‌ها در صحنه‌ای از فیلم باید چه شکلی باشد موسیقی مناسب تری را تولید کرد حتی با استفاده از پالس‌های عصبی افراد می‌توان موسیقی مناسب برای هر فرد را تولید کرده و برای بیماری‌های خاصی از آن استفاده کرد.

## کتاب نامه

- [1] Lesiuk, Teresa. The effect of music listening on work performance. *Psychology of music*, 33(2):173–191, 2005.
- [2] Siedliecki, Sandra L and Good, Marion. Effect of music on power, pain, depression and disability. *Journal of advanced nursing*, 54(5):553–562, 2006.
- [3] Roballey, Thomas C, McGreevy, Colleen, Rongo, Richard R, Schwantes, Michelle L, Steger, Peter J, Wininger, Marie A, and Gardner, Elizabeth B. The effect of music on eating behavior. *Bulletin of the Psychonomic Society*, 23(3):221–222, 1985.
- [4] Anshel, Mark H and Marisi, Dan Q. Effect of music and rhythm on physical performance. *Research Quarterly. American Alliance for Health, Physical Education and Recreation*, 49(2):109–113, 1978.
- [5] Pelletier, Cori L. The effect of music on decreasing arousal due to stress: A meta-analysis. *Journal of music therapy*, 41(3):192–214, 2004.
- [6] Hee Park, Hyun, Kwan Park, Jai, and Ok Jeon, Jung. Attributes of background music and consumers' responses to tv commercials: The moderating effect of consumer involvement. *International Journal of Advertising*, 33(4):767–784, 2014.
- [7] Hoeberichts, Nina. Music and advertising: The effect of music in television commercials on consumer attitudes. *Bachelor Thesis*, 2012.
- [8] Costabile, Kristi A and Terman, Amanda W. Effects of film music on psycholog-

- ical transportation and narrative persuasion. *Basic and Applied Social Psychology*, 35(3):316–324, 2013.
- [9] Pavlović, Ivanka and Marković, Slobodan. The effect of music background on the emotional appraisal of film sequences. *Psihologija*, 44(1):71–91, 2011.
- [10] Li, Xiao-Mei, Zhou, Kai-Na, Yan, Hong, Wang, Duo-Lao, and Zhang, Yin-Ping. Effects of music therapy on anxiety of patients with breast cancer after radical mastectomy: a randomized clinical trial. *Journal of advanced nursing*, 68(5):1145–1155, 2012.
- [11] Yamashita, Kaoru, Kibe, Toshiro, Ohno, Sachi, Kohjitani, Atsushi, and Sugimura, Mitsutaka. The effects of music listening during extraction of the impacted mandibular third molar on the autonomic nervous system and psychological state. *Journal of Oral and Maxillofacial Surgery*, 77(6):1153–e1, 2019.
- [12] Pothoulaki, MacDonald, Macdonald, Raymond AR, Flowers, Paul, Stamatakis, E, Filiopoulos, V, Stamatiadis, D, and Stathakis, Ch P. An investigation of the effects of music on anxiety and pain perception in patients undergoing haemodialysis treatment. *Journal of Health Psychology*, 13(7):912–920, 2008.
- [13] Ongsulee, Pariwat. Artificial intelligence, machine learning and deep learning. in *2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE)*, pp. 1–6. IEEE, 2017.
- [14] Hewahi, Nabil, AlSaigal, Salman, and AlJanahi, Sulaiman. Generation of music pieces using machine learning: long short-term memory neural networks approach. *Arab Journal of Basic and Applied Sciences*, 26(1):397–413, 2019.