الجمهورية العربية السورية

وزارة التعليم العالي والبحث العلمي

جامعة دمشق

كلية الهندسة الميكانيكية والكهربائية

قسم هندسة الإلكترونيات والاتصالات

استرجاع المعلومات الموسيقية باستخدام معالجة الإشارة الرقمية والتعلم العميق

رسالة مقدمة لنيل درجة الماجستير في هندسة الإلكترونيات التطبيقية

إعداد

م. إبراهيم وهبي

إشراف

د.م. حسين طياوي بحبوح

العام الدراسي

2023-2022

صفحة قرار الحكم مع توقيع اللجنة

تصريح خطي

أنا الموقع أدناه أصرح بعلمي الكامل وقبولي:

1. أن كل ما ينتج عن البحث والأطروحة هو ملكية فكرية ومالية بالتشارك مع جامعة دمشق، وأنني ألتزم بأخذ موافقة الجامعة في حال رغبتي بنشر البحث أو الأطروحة أو جزء منها نصاً أو مضموناً خارج إطار الجامعة (من دور نشر أو مكتبات أو مواقع إلكترونية وغيرها من وسائل النشر).
2. أنه لا يوجد أي جزء من هذه الأطروحة تم اقتباسه من عمل علمي آخر أو أنجز للحصول على شهادة أخرى في جامعة دمشق أو أية جامعة أو معهد تعليمي داخل الجمهورية العربية السورية أو خارجها.

الاسم والتوقيع

الإهداء

إلى وطني الغالي، سوريا الحبيبة

إلى الرجل الذي دعا الله بالتوفيق لي – أبي الغالي

إلى من مهدت لي درب النجاح – أمي الحبيبة

إلى من أرشدني في كل خطوة – د. حسين بحبوح

إلى كل محب وصديق

شكر وتقدير

قال تعالى: **{يَرْفَعِ اللَّهُ الَّذِينَ آمَنُوا مِنْكُمْ وَالَّذِينَ أُوتُوا الْعِلْمَ دَرَجاتٍ وَاللَّهُ بِمَا تَعْمَلُونَ خَبِيرٌ}** صدق الله العظيم

الحمد لله رب العالمين الذي أكرمني بهذا العمل المتواضع ويسّر لي مصاعبه وعقباته جميعها وأرجو من الله أن يكون عملي هذا وسيلة تقربني إليه وتجعله راضيا عني.

إلى أمي وأبي ... إلى القناديل المضيئة في حياتي التي أكرمني الله بها ... إلى من كان دعاؤهم سرّ نجاحي ... إلى من ستبقى كلماتهم نجوم أهتدي بها في الأمس واليوم وأبدا.

إلى أخي وأصدقائي والقلوب الطاهرة الرقيقة وكل محب وداع لي بالخير.

كما أتوجه بالشكر الجزيل إلى كل من أسهم في مسيرتي الدراسية وإلى كل من دعمني وساندني في إنجاز هذا البحث.

وأخص بعميق الشكر والامتنان السادة أعضاء لجنة الحكم الذين كلفوا أنفسهم عناء قراءة هذا العمل وستكون توجيهاتهم لي دافعاً لتجويد العمل والارتقاء به.

وأقدم أسمى آيات الشكر والامتنان إلى أساتذتي الأفاضل جميعهم في قسم الهندسة الإلكترونية في جامعتي دمشق وتشرين وأخص بالذكر الدكتور حسين طياوي بحبوح الذي أشرف على هذا البحث.

فهرس المحتويات

صفحة قرار الحكم مع توقيع اللجنة 2

تصريح خطي..... 3

الإهداء..... 4

شكر وتقدير..... 5

فهرس المحتويات.. 6

فهرس الأشكال التوضيحية 9

فهرس الجداول 10

قائمة المصطلحات والاختصارات..... 11

المستخلص العربي..... 12

المستخلص الإنكليزي Abstract..... 14

1-الفصل الأول: الإطار العام للبحث 17

1-1- المقدمة: 18

1-2 بعض الدراسات السابقة: 19

1-3- مشكلة البحث: 23

1-4 هدف البحث: 23

1-5 أهمية البحث: 23

1-6 مسوغات البحث: 24

1-7- محدّدات البحث: 24

1-8 حدود البحث: 24

1-9- منهج البحث وأدواته: 25

1-10- بنية الرسالة:..... 25

2-الفصل الثاني: الصوت والموسيقا ومعالجة الإشارة الرقمية.....**29**

2-1 الصوت:..... 30
2-2 الموسيقا: 30
2-3 واجهة الآلات الموسيقية الرقمية Musical Instrument Digital Interface:..... 32
2-4 أطوار النغمة الموسيقية: 35
2-5 معالجة الإشارة الرقمية Digital Signal Processing (DSP): 36
2-5-1 تحويل فوريه المتقطع Discreet Fourier Transform (DFT): 37
2-5-2 تحويل فوريه قصير الزمن Short Time Fourier Transform (STFT): 38
2-5-3 تحويل عامل الجودة الثابت Constant-Q-Transform (CQT): 40
2-6-1 اكتشاف النغمات الأحادية Mono-pitch Detection:..... 43
2-6-2 اكتشاف النغمات المتعددة Multi-pitch Detection:..... 43
3- الفصل الثالث: تعلم الآلة والشبكات العصبونية 45
3-1 تعلم الآلة Machin Learning:..... 46
3-1-1 شبكات الذكاء الصنعي:..... 46
3-1-2 الانتشار الأمامي Forward Propagation :..... 47
3-1-3 توابع التنشيط Activation Functions: 48
3-1-4 توقع النتائج:..... 48
3-1-5 الانتشار الخلفي Back Word Propagation:..... 49
3-2 الشبكات العصبونية الالتفافية Convolutional Neural Networks (CNN):..... 50
3-3 التعلم العميق Deep Learning: 53
4- الفصل الرابع: جمع بيانات التدريب والمعالجة المسبقة للعينات الموسيقية..... 55
4-1 مقدمة: 56
4-2 قاعدة بيانات التدريب Training Dataset: 56
4-2-1 قاعدة بيانات MIDI Aligned Piano Sounds (MAPS):..... 57

4-2-2 إنشاء قاعدة بيانات للموسيقا الشرقية: 57

4-2-2-1 تعديل زمن امتداد النغمة الموسيقية: 58

4-2-2-3 تحويل مقاطع MIDI إلى ملفات صوتية: 63

4-2-3 تحويل صيغة MIDI لمصفوفة ثنائية: 63

4-3 معالجة ملفات WAV: 65

4-3-1 تنفيذ تحليل CQT على الملفات الصوتية: 65

4-3-2 معالجة الطيف الناتج عن تحليل CQT: 67

5- الفصل الخامس: تصميم النموذج والنتائج 72

5-1 بناء النموذج: 73

5-1-1 بناء الشبكة العصبونية: 73

5-1-2 توابع تنشيط الشبكة العصبونية: 76

5-1-3 بارمترات الشبكة العصبونية: 77

5-1-4 تحويل خرج الشبكة العصبونية إلى ملف MIDI: 78

5-2 تقييم الأداء: 79

5-2-1 معايير التقييم: 79

5-2-2آلية التقييم: 81

5-3 النتائج: 82

5-4 مقارنة النتائج مع الدراسات السابقة: 85

5-5 الخاتمة Conclusion: 87

الآفاق المستقبلية 88

المراجع 89

فهرس الأشكال التوضيحية

الشكل(1-1) البنية العامة للنظام في الدراسة [9]..... 20
الشكل (1-2) البنية العامة للنموذج المستخدم في الدراسة [10]..... 21
الشكل (1-3) مخطط يوضح بنية الشبكة في الدراسة [8] 22
الشكل (1-4) المخطط الصندوقي للمنهجية المتبعة 26
الشكل (2-1) الفرق بين ربع العلامة ونصف العلامة 30
الشكل (2-2) مخطط يبين بداية ونهاية وشدة العلامات الموسيقية في جزء من مقطوعة موسيقية 30
الشكل (2-3) تمثيل ملف MIDI الزمني على شكل لفات بيانو (Piano roll) 31
الشكل (2-4) محددات صيغة MIDI الرقمية 33
الشكل (2-5) الأطوار المختلفة للنغمة الموسيقية..... 35
الشكل (2-6) تحويل فورييه السريع لإشارة تحوي 10، 20، 30، 40، 50 هرتز 37
الشكل (2-7) تمثيل لسلم الميجور بصيغة MIDI ولفات بيانو واستجابة ترددية..... 39
الشكل (2-8) مقارنة بين تحويل STFT وتحويل CQT لسلم الدو ميجور الموسيقي 40
الشكل (2-9) نافذة Hann واستجابتها الترددية 41
الشكل (2-10) التوافقيات الترددية لعلامة دو في الأوكتافين الخامس والسادس..... 43
الشكل (3-1) مثال عن نموذج شبكة عصبونية كاملة الاتصال بطبقتين مخفيتين.....46
الشكل (3-2) توابع التنشيط الأكثر شيوعا 47
الشكل (3-3) بنية الشبكة العصبونية الالتفافية..... 51
الشكل (3-4) المميزات المكتشفة من قبل الشبكة العصبونية في نظام لتصنيف الوجوه 52
الشكل (4-1) حدث امتداد النغمة الموسيقية 57
الشكل (4-2) خوارزمية تعديل طول العلامة المستعملة..... 42
الشكل (4-3) واجهة برنامج تعديل قاعدة بيانات MIDI..... 60

الشكل (4–4) خوارزمية تعديل المقام الموسيقي 61
الشكل (4–5) آلية تحويل ملف MIDI إلى مصفوفتي بداية وامتداد النغمة..... 63
الشكل (4–6) الطيف الناتج عن تحويل CQT لـ 12 ثانية لمقطوعة موسيقية..... 65
الشكل (4–7) الطيف الناتج عن تحويل CQT قبل حساب منقول المصفوفة 68
الشكل (4–8) مثال عن الطيف الناتج عن تحويل CQT بعد حساب منقول المصفوفة وتجميع الإطارات..... 69
الشكل (5–1) البنية العامة للنموذج المصمم.....72
الشكل (5–2) بنية الشبكة العصبونية من الدخل حتى الخرج مع أحجام الطبقات 73
الشكل (5–3) بنية كاشف امتداد النغمة الموسيقية بدون طبقات Dropout و Batch normalization 75
شكل (5–4) واجهة برنامج تحويل خرج الشبكة العصبونية إلى ملف MIDI 78
شكل (5–5) لفات البيانو piano roll لملف MIDI المستخرج من الشبكة العصبونية لبداية النغمة فقط..... 82
شكل (5–6) لفات البيانو لملف MIDI المستخرج لكل من بداية النغمة وامتداد النغمة 83

فهرس الجداول

جدول (4–1) المحددات المستخدمة في تحليل CQT 65
جدول (5–1) توابع التنشيط المستخدمة في الشبكة العصبونية 76
جدول (5–2) المحددات المستخدمة في تصميم الشبكة العصبونية 77
جدول (5–3) جدول الحقيقة للشبكة العصبونية..... 79
جدول (5–4) مثال لصيغة الدخل لمكتبة MIR_EVAL 81
جدول (5–5) نتائج التقييم لكل من الإحكام (P)، والحساسية (R)، ودرجة F1 81
جدول (5–6) مقارنة الإحكام (P)، والحساسية (R)، ودرجة F1 مع الدراسات التي حسبت خطأ ربع العلامة.... 84
جدول (5–7) مقارنة الإحكام (P)، والحساسية (R)، ودرجة F1 مع الدراسات التي لم تحسب خطأ ربع العلامة. 84

قائمة المصطلحات والاختصارات

المصطلح باللغة الإنكليزية	الاختصار	المرادف باللغة العربية
Automatic Music Transcription	AMT	كتابة الموسيقا الآلي
Central Processing Units	CPUs	وحدات المعالجة المركزية
Constant-Q-Transform	CQT	تحويل عامل الجودة الثابت
Convolutional neural networks	CNN	الشبكات العصبونية الالتفافية
Digital Signal Processing	DSP	معالجة الإشارة الرقمية
Discreet Fourier Transform	DFT	تحويل فورييه المتقطع
Dynamic Bayesian Networks	DMN	شبكات بايسيان الديناميكية
False Negative	FN	سلب كاذب
False Positive	FP	إيجاب كاذب
Fast Fourier Transform	FFT	تحويل فوريه السريع
Fundamental Frequency	F0	التردد الأساسي
Graphic processing units	GPU	وحدات معالجة البيانيات
Hidden Markov Model	HMM	نموذج ماركوف المخفي
Long Short Term Memory	LSTM	الذاكرة الطويلة قصيرة المدى
Musical Information Retrieval	MIR	استرجاع المعلومات الموسيقية
Musical Instrument Digital Interface	MIDI	الواجهة الرقمية للآلات الموسيقية
Nonnegative Matrix Factorization	NMF	تحويل المصفوفات الغير سالبة إلى عوامل
probabilistic latent component analysis	PLCA	تحليل المكونات الكامنة الاحتمالية
Pulse Code Modulation	PCM	تعديل رمز النبضة
Recurrent Neural Network	RNN	الشبكات العصبونية التكرارية
Short Time Fourier Transform	STFT	تحويل فورييه قصير الأمد
True Negative	TN	سلب حقيقي
True Positive	TP	إيجاب حقيقي
Waveform Audio File Format	WAV	ملف صوت شكل الموجة

المستخلص العربي

خلفية البحث:

تعد عملية استرجاع المعلومات الموسيقية عملية معقدة ولا سيّما في المقطوعات الموسيقية متعددة النغمات، فضلاً عن ذلك فإنّ معظم الدراسات السابقة لم تأخذ خطأ ربع العلامة الموسيقية في الحسبان علماً بأنّ ربع العلامة الموسيقية تعد السمة الأهم في الموسيقا العربية.

هدف البحث:

بناء نموذج كامل لتتبع النغمات المتعددة للموسيقا العربية والغربية على حد سواء، الذي يتميز بقدرته على اكتشاف بدايات النغمات وامتدادها لمقطوعات البيانو الموسيقية.

المواد والطرائق:

يجمع هذا النموذج كل من تقانات معالجة الإشارة الرقمية والتعلم العميق لزيادة دقة النتائج. بداية يُطبّق تحويل عامل الجودة الثابت على ملف الدخل الصوتي، وتمرير نتيجة التحويل للمرحلة الثانية التي تحتوي على شبكتين عصبونيتين التفافيتين. صممت الشبكة الأولى لاكتشاف بداية النغمة الموسيقية، في حين صممت الشبكة الثانية لاكتشاف امتداد هذه النغمة.

النتائج العملية:

يُعالج خرج الشبكتين العصبونيتين آنفتي الذكر وتوليد ملف بامتداد MIDI للملف الصوتي المدخل وذلك بدرجة F1 تبلغ قيمتها 75.28% لبداية النغمة الموسيقية و 47.37% للإطار كاملا.

الاستنتاجات من وجهة نظر الباحث:

اعتمادا على النتائج التي خرج بها هذا البحث تبيّن إمكانية استخدام الشبكات العصبونية الالتفافية وتحويل عامل الجودة الثابت لاكتشاف بداية ونهاية العلامة الموسيقية وذلك عندما يكون خطأ ربع العلامة الموسيقية محسوبا؛ ويمكن استخدام النموذج المبني في معاهد تعليم الموسيقى واستخراج النوط الموسيقية.

الكلمات المفتاحية:

كتابة الموسيقا، استرجاع المعلومات الموسيقية، تخمين التردد f0، اكتشاف بداية النغمة، موسيقا شرقية، التعلم العميق، تحويل عامل الجودة الثابت، الشبكات العصبونية الالتفافية.

Abstract المستخلص الإنكليزي

Research Background:

Musical information retrieval is considered a complex operation, especially in polyphonic music tracks. However, most of the previous studies haven't taken into account quarter notes detection which is considered the most important feature in Arabic and Turkish music.

Aim of Research:

In this thesis, out Aim was to build a complete polyphonic pitch tracking system for Arabic, and Western music which is able to detect notes and framewise from piano music tracks.

Materials And Methods:

This model combines both signal processing and deep learning techniques to provide more accurate results. In the first stage, constant Q Transformation is applied to the input musical track and then passed

to the next stage which consists of two convolutional neural networks (CNNs). The first network is designed to detect notes onset and the other network is for framewise pitch detection.

Practical Results:

The output of the neural network is processed and generates a MIDI file for the input music track with an F1 score of 75.28% for onset detection and 47.37% for the framewise.

Conclusion from Researcher Point of View:

Based on the results of this research, it has been shown that it is possible to use convolutional neural networks and a constant quality factor conversion to detect the beginning and end of a musical note when the quarter note error is detriment. The developed model can be used in music education institutions to extract musical scores.

Keywords:

Music transcription, f0 estimation, onset detection, eastern music, deep learning, CQT, CNN, pitch, musical information retrieval

الفصل الأول

الإطار العام للبحث

Research General Framework

1-1-المقدمة:

تعرّف عملية استرجاع المعلومات الموسيقية Musical Information Retrieval (MIR) على أنها عملية استخلاص المعلومات الرقمية من الموسيقا المسموعة والملفات الصوتية [1]، وتحظى هذه العملية باهتمام متزايد من الباحثين مع العديد من التطبيقات المستخدمة في الحياة العملية. يتضمن هذا المجال عددا من العمليات الجزئية ولعل أبرزها هو إعادة كتابة الموسيقا بشكل آلي Automatic Music Transcription (AMT) وهي عملية تحويل الإشارة الصوتية إلى رموز مكتوبة كالنوطة الموسيقية، أو إلى صيغة يفهمها الحاسب كالواجهة الرقمية للآلات الموسيقية Musical Instrument Digital Interface (MIDI). تتلخص المشكلة الأساسية في هذا المجال باكتشاف مميزات النغمة وهي:

- زمن بداية النغمة.
- زمن نهاية النغمة.
- شدة النغمة.
- التردد الأساسي Frequency (F0) .
- طابع الصوت Tempo (للتمييز بين نغمتين لهما نفس التردد والشدة).

فضلاً عن اكتشاف الآلات الموسيقية المستخدمة، واكتشاف التوافقيات الترددية، والأنماط الصوتية المتكررة. تعد هذه العمليات صعبة حتى على العازفين المتمكنين، وتزداد صعوبة كلما زاد عدد النغمات المعزوفة في وقت واحد وكلما زاد عدد الآلات الموسيقية المستخدمة مما يشكل تحدياً كبيراً للباحثين في هذا المجال. بخلاف الموسيقا الغربية التي تعتمد على وجود نصف علامة موسيقية بين كل نغمتين متتاليتين فإنّ الموسيقا العربية تعتمد على وجود ربع علامة موسيقية ما يزيد عدد العلامات الموسيقية في كل أوكتاف (يعرف موسيقيا على أنه مجموعة العلامات الموسيقة من علامة دو حتى علامة السي الموسيقية) من 12 حتى 24 علامة. ومن الجدير

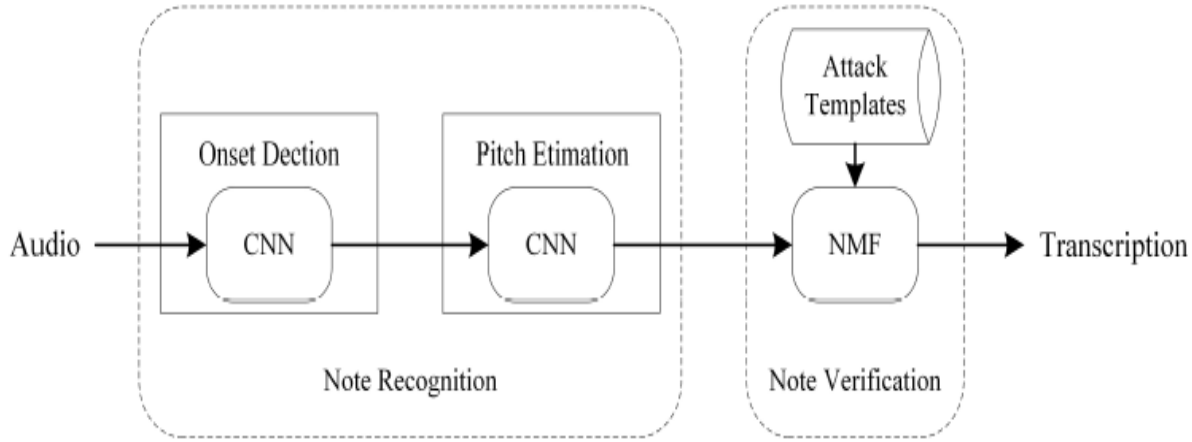
بالذكر أنّ الغالبية العظمى من الدراسات السابقة استهدفت الموسيقا ذات الطابع الغربي (الأوروبي) [2]، كما أنّ آلية تقييم النتائج المستخدمة في هذه الدراسات لم تأخذ خطأ ربع العلامة الموسيقية بالحسبان [3]. إنّ الدراسات السابقة ولا سيّما التي أخذت الموسيقا الشرقية بالحسبان استخدمت أساليب معالجة الإشارة مثل تحويل فوريه السريع Fast Fourier Transform (FFT) [4] وتقنيات تحليل المصفوفات غير السالبة إلى عوامل Nonnegative matrix factorization (NMF) [5]. ولجأت بعض الأبحاث إلى توقع النغمات بملاحقة الاستطاعات الأعظمية للترددات السائدة باستخدام شبكات بايسيان الديناميكية dynamic Bayesian networks (DBN) [6] أو باستخدام نموذج ماركوف المخفي hidden Markov Model (HMM) [7] في حين كان التوجه الجديد والذي حقق حالة المعرفة في هذا المجال يعتمد على استخدام تقنيات التعلم العميق مثل الشبكات العصبونية الالتفافية Convolutional neural networks (CNN) (وهي المستخدمة في هذا البحث) والشبكات العصبونية التكرارية recurrent neural network (RNN) أو مزيجا بين التقنيتين آنفتي الذكر[8] .

1-2 بعض الدراسات السابقة:

شهد مجال استرجاع المعلومات الموسيقية اهتماماً ونمواً متزايداً في المدّة الأخيرة ولا سيّما في مجال كتابة الموسيقا الآلية ونذكر بعض الدراسات السابقة التي تناولت هذا الموضوع:

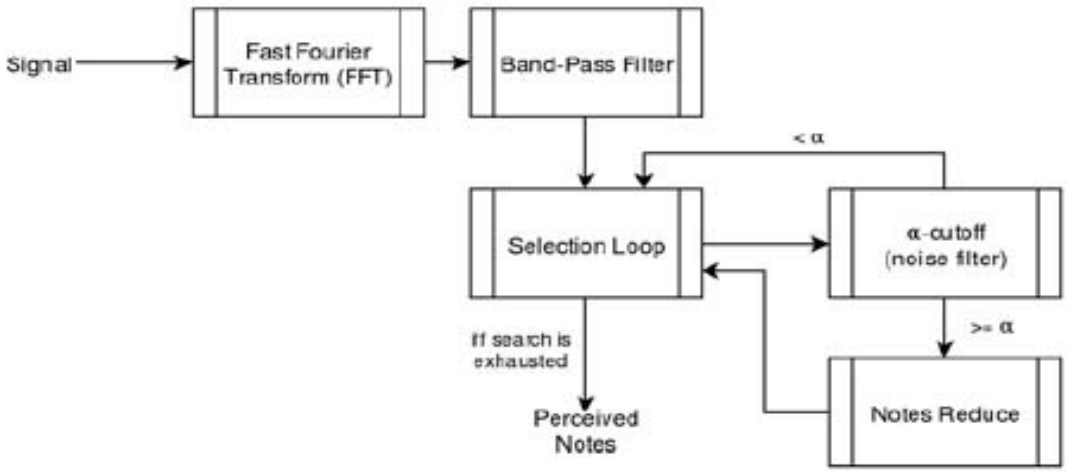
في الدراسة [2] استُخدِم تحويل عامل الجودة الثابت وتحليل المكونات الكامنة الاحتمالية probabilistic latent component analysis (PLCA) لاكتشاف بداية النغمة الموسيقية والتردد الخاص بها آخذة بالحسبان ربع العلامة الموسيقية لآلات موسيقية شرقية وغربية عدّة وحققت درجة F1 score بقيمة 51.91% في اكتشاف بداية النغمة ضمن زمن إطار 100 ms (سنة النشر 2015 ضمن The Journal of the Acoustical Society of America).

أما في الدراسة [9] فقد استُخدم نهج ذو مرحلتين لاكتشاف بداية العلامة الموسيقية وتحديد هذه العلامة لآلة بيانو محددة مسبقا. تتكون المرحلة الأولى من شبكتين عصبونيتين التفافيتين تقوم الشبكة الأولى بتحديد بداية العلامة الموسيقية بينما تقوم الشبكة الثانية بتحديد العلامة الموسيقية، لتُمرّر النتيجة إلى المرحلة الثانية التي تتكون من نموذج لتوليد طور هجوم العلامة المكتشفة نفسها (أول أطوار النغمة الموسيقية) Attack template ويقارن مع الملف الأصلي للتأكد من صحة النتيجة. حققت الدراسة F1 score بقيمة 80.23% ولكن الدراسة أجريت على صوت بيانو واحد فقط ما يؤثر على موضوعية النتائج كما أنّها لم تأخذ خطأ ربع العلامة بالحسبان. (سنة النشر 2017 ضمن MDPI). يوضح الشكل (1-1) البنية العامة للنظام في الدراسة آنفة الذكر.



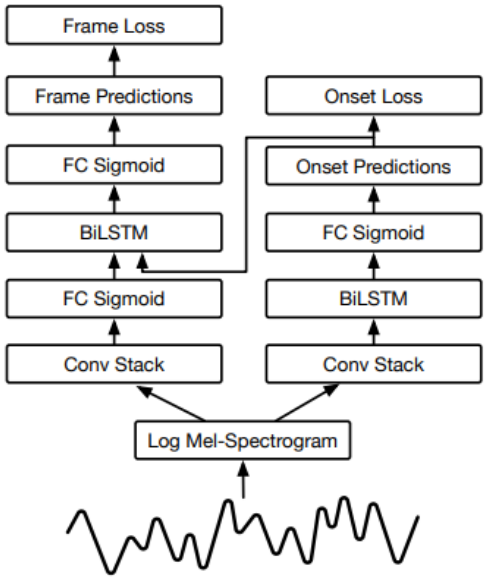
الشكل (1-1) البنية العامة للنظام في الدراسة [9]

قامت الدراسة [10] باستخدام تحويل فوريه السريع ومرشحات تمرير الحزمة ومرشحات الضجيج لاستخلاص العلامات الموسيقية في الوقت الحقيقي وحققت F1 score بقيمة 84.42% والذي يعد جيدا نظرا لزمن المعالجة القصير إلا أنها لم تأخذ خطأ الأوكتاف في الحسبان والذي يعد الخطأ الأكثر شيوعا (سنة النشر 2018 ضمن IEEE). يوضح الشكل (1–2) البنية العامة للنموذج المستخدم في الدراسة آنفة الذكر.



الشكل (1–2) البنية العامة للنموذج المستخدم في الدراسة [10]

قدمت هذه الدراسة [8] حالة المعرفة بالنسبة لاسترجاع المعلومات الموسيقية وفق ما ذكرته الدراسة وهي من إعداد شركة غوغل استخدمت نوعين من الشبكات وهي شبكة عصبونية التفافية وشبكة ذات ذاكرة طويلة وقصيرة المدى Long Short Term Memory (LSTM). يوضح الشكل (1–3) بنية الشبكة العصبونية في الدراسة آنفة الذكر.



الشكل (1-3) مخطط يوضح بنية الشبكة في الدراسة [8]

واستُخدمت للتدريب قاعدة بيانات "MAPS" وتحوي مقطوعات موسيقية للبيانو بحجم 31 غيغا بايت ومقاطع صوتية بصيغة WAV وملف MIDI المقابل لها (ملف العلامات الموسيقية) وحققت درجة F1 بقيمة 82.29% في اكتشاف بداية النغمة مع العلامة الموافقة لها ودرجة 50.22% عند تضمين نهاية النغمة الموسيقية والإطار كاملا. لم تأخذ هذه الدراسة خطأ ربع العلامة بالحسبان. (سنة النشر 2018 ضمن arXiv).

قامت الدراسة [11] بتصميم نظام كامل لاسترجاع الموسيقا إلى نغمات موسيقية مستخدمة 6 شبكات عصبونية التفافية CNN 3 منها لتحديد البنى الأساسية للنغمة (بداية ونهاية وتردد النغمة) بينما شكلت الشبكات العصبونية الباقية مراحل بينية يشكل خرجها دخلا للشبكات آنفة الذكر وقد حسنت من دقة اكتشاف الإطار عند أخذ نهاية العلامة بالحسبان لتصبح قيمة F1 score 58% في حين كانت قيمة اكتشاف تردد النغمة مع بداية العلامة الموسيقية 78.40% ومن سلبيات هذه الدراسة أنها لم تأخذ خطأ ربع العلامة بعين الاعتبار.

1-3-مشكلة البحث:

يعد تحويل الموسيقا المسموعة إلى مخطط موسيقي أو نوطه موسيقية من العمليات المهمة والمعقدة في المجال الموسيقي. فعلى الرغم من أنّ العديد من الموسيقيين المحترفين يمارسون العزف اعتمادا على حاسة السمع إلا أنّها عملية صعبة وغير دقيقة وتمثل تحدياً حقيقياً لا سيّما بالنسبة للمبتدئين. كما أنّ الدراسات السابقة تركز على الموسيقا الغربية فقط إذ أنّ بنية الموسيقا الشرقية تختلف اختلافاً كبيراً عن الموسيقا الغربية.

1-4 هدف البحث:

1- دراسة الحلول الحالية في مجال استرجاع المعلومات الموسيقية ودراسة إمكانية تطبيقها على الموسيقا الشرقية.

2- اقتراح طريقة واحدة أو أكثر لاسترجاع المعلومات الموسيقية في الموسيقا الشرقية باستخدام التعلم العميق والتحويلات والمرشحات المناسبة.

3- تطوير النهج المقترح وتقييم أدائه مقارنة بالحلول الحالية بناء على قواعد البيانات المتاحة.

1-5 أهمية البحث:

1- تعد الشريحة المستهدفة في هذه الدراسة كبيرة بوصفها عدداً.

2- قلة الأبحاث في هذا المجال ولا سيّما الموسيقا الشرقية.

3- حل مشكلة صعبة لدى العازفين ولا سيّما المبتدئين منهم.

1-6 مسوغات البحث:

إنّ قلة الدراسات والأدوات التي استهدفت مجال استرجاع المعلومات الموسيقية في الموسيقة الشرقية، خلق حاجة لدى العازفين ولا سيّما المبتدئين لإيجاد طريقة سهلة وسريعة تحول الموسيقا المسموعة إلى تدوين موسيقي.

1-7- محدّدات البحث:

1- محدودية قدرة المعالجة الحاسوبية نظرا لما تتطلبه الشبكات العصبونية ولا سيّما شبكات التعلم العميق من كمية معالجة كبيرة للبيانات.

2- عدم توفر عينات التدريب الكافية للموسيقا الشرقية في الإنترنت.

1-8 حدود البحث:

- **الحد المكاني:** جرى تنفيذ هذا البحث في قسم هندسة الإلكترونات والاتصالات في كلية الهمك بجامعة دمشق.
- **الحد التقاني:** جرى تنفيذ النظام المقترح على جهاز حاسب شخصي ذو معالج i5 من الجيل الخامس وذاكرة وصول عشوائية ذات حجم 12Gb واستخدام Google colab لتدريب الشبكة العصبونية.
- **الحد الزمني:** من الشهر الثاني عام 2021 حتى الشهر السابع عام 2023.
- **الحد العلمي:** جرى استخدام كل من المنهج التجريبي والاستقرائي لتدريب الشبكة العصبونية والحصول على قيم أفضل لمحددات الشبكة كما تم اعتماد المنهج الاستقرائي لتوليد عينات تدريب الموسيقى الشرقية وذلك بتحويل عينات الموسيقى الغربية لموسيقى شرقية.

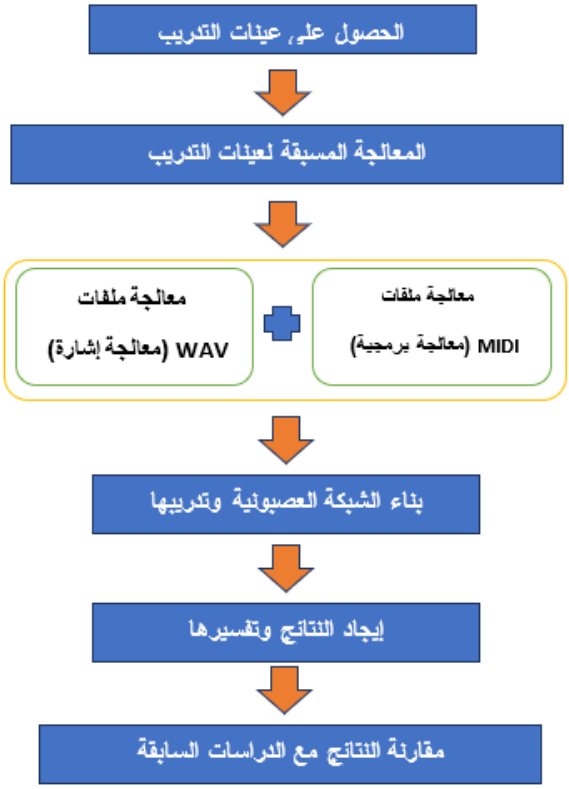
1-9- منهج البحث وأدواته:

تمثل منهجية البحث الموضحة في الشكل (1-4) مجموعة الخطوات المعتمدة للتأكد من سلامة تصميم النموذج المقترح وتقييم أدائه، كما أن دراسات الباحثين السابقة في هذا المجال تتيح لنا الاطلاع على طرائقهم وأساليبهم، وتحدد آلية ومنهجية للتأكد من مستوى أداء النظام المقترح، من ثمّ فإنّنا سنتبع الطرائق والمعايير الموجودة في الدراسات السابقة لبناء النظام وتقييم أدائه وتتلخص بالخطوات التالية:

1- مراجعة الأدبيات العلمية في هذا المجال.

2- اقتراح طريقة لتحويل المقطوعات الموسيقية إلى ملف MIDI باستخدام تحويل عامل الجودة الثابت CQT Constant-Q-Transform وشبكات CNN.

- الحصول على عينات التدريب للموسيقا الشرقية والغربية بكلّ من صيغتي: ملف صوت شكل الموجة Waveform Audio File Format (WAV) وMIDI.
- المعالجة المسبقة لعينات التدريب كي تلائم بنية الموسيقا الشرقية والشبكة العصبونية (معالجة ملفات MIDI معالجة برمجية ومعالجة إشارة لملفات WAV).
- بناء الشبكة العصبونية وتدريبها.
- إيجاد النتائج وفقا للمعايير المتبعة في الدراسات السابقة وتحليلها ومقارنتها مع الدراسات السابقة.



الشكل (1-4) المخطط الصندوقي للمنهجية المتبعة

جرى استخدام الأدوات البرمجية الآتية:

1- لغة برمجة .Net 4.5 c# (لبناء برامج تعديل ملفات MIDI وتحويل خرج الشبكة العصبونية إلى ملفMIDI بواسطة محرر Visual Studio)

2- برنامج Sonic Visualizer V4.1 (دراسة مرجعية للتحويلات المستخدمة في معالجة الإشارة الصوتية)

3- برنامج Cubase 8.0 (تعديل ملفات MIDI).

4- برنامج Reaper 6.59 (تعديل ملفات MIDI وتوليد ملفات صوتية منها).

5- موقع Google colab (تدريب الشبكة العصبونية).

6- لغة برمجة Python 3.9 64-bit (معالجة الإشارة وبناء الشبكة العصبونية).

- مكتبة TensorFlow Library 2.7.0 (بناء الشبكة العصبونية).
- مكتبة Keras API 2.7.0 (بناء الشبكة العصبونية).
- مكتبة Librosa(تطبيق تحويلات معالجة الإشارة على الملفات الصوتية).

كما جرى استخدام العتاديات الآتية:

1- حاسب محمول ذو مواصفات (CPU: i5-5300U , Ram 12 GB DDR3).

2- وصلة MIDI.

1-10 بنية الرسالة:

رُتّبت الرسالة في خمسة فصول:

- عرض الفصل الأول الإطار العام للبحث والذي يتضمن ما يأتي:
  - مقدمة.
  - بعض الدراسات السابقة.
  - مشكلة البحث.
  - هدف البحث.
  - أهمية البحث.
  - مسوّغات البحث.
  - محدّدات البحث.
  - حدود البحث.
  - منهج البحث وأدواته.
  - بنية الرسالة.
- يتضمن الفصل الثاني بعض أساسيات الصوت والموسيقا وتقنيات معالجة الإشارة الرقمية المستخدمة في البحث.

- يتحدث الفصل الثالث عن تعلم الآلة والشبكات العصبونية ومبدأ عملها وميزات كل نوع منها.
- يناقش الفصل الرابع عملية جمع عينات التدريب والمعالجة المسبقة لها وآلية حل مشكلة عدم توفر هذه العينات من مصادر سابقة.
- يتناول الفصل الخامس عملية تصميم النموذج والشبكة العصبونية والنتائج المستخلصة من الدراسة.

وأخيرا خلاصة العمل والأعمال المستقبلية والمراجع العلمية المستخدمة.

الفصل الثاني

الصوت والموسيقا ومعالجة الإشارة الرقمية

Sound, Music and Digital Signal Processing

2-1 الصوت:

يعرف الصوت من منظور المستمع على أنه تفاعل الأذن مع مجال محدد من الاهتزازات وله محددان أساسيان هما شدة الصوت Volume والنغمة Pitch.

تحدد النغمة ما إذا كان الصوت يوصف نغمةً عاليةً أو منخفضةً مقارنةً بباقي النغمات بالنسبة للمستمع. وترتبط النغمة بعدد الاهتزازات المتولدة لِمدّة زمنية معينة، ويطلق على عدد هذه الاهتزازات في الثانية الواحدة بالتردد فكلما زاد التردد أصبحت النغمة الموسيقية المسموعة أعلى ويقاس التردد عادة بوحدة الهرتز (Hz).

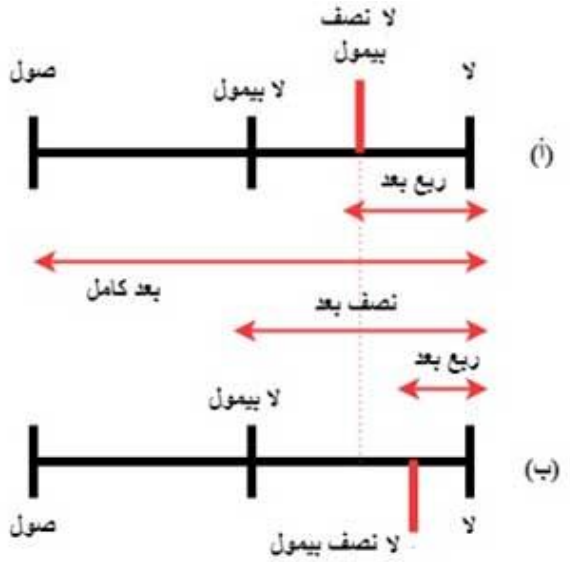
تصدر الآلات الموسيقية عدداً من التوافقيات الترددية عند عزف علامة موسيقية مفردة إلا أنَّ النغمة المحسوسة غالباً ما تكون ذات التردد الأدنى في هذه التوافقيات وهو ما يدعى بالتردد الأساسي fundamental frequency (F0) في حين تكون التوافقيات مضاعفات لهذا التردد.

تعرف شدة الصوت على أنّها مستوى استطاعة الإشارة الصوتية المسموعة (مطال الموجة الصوتية) وعادة ما تقاس بالديسبل (dB). كما أنَّ طابع الصوت Timber يعد مفهوماً آخرا مهماً يسمح لنا بالتمييز بين مصادر الصوت المختلفة التي تصدر نغمات متماثلة بشدة صوت متساوية، إذ أنّ كل آلة موسيقية تتميز بطابع صوتي مختلف عن الأخرى، ويعد هذا الطابع مرتبطا بهوية الآلة وبأسلوب العزف عليها، وقد نجد في العديد من الحالات أن آلتين موسيقيتين من نفس النوع قد تحملان طابعاً صوتياً مختلفاً، وذلك بسبب اختلاف نسب التوافقيات الترددية وانحراف العلامة الموسيقية عن التردد الصحيح لها بمقدار بسيط وسرعة تخامد الإشارة الصوتية.

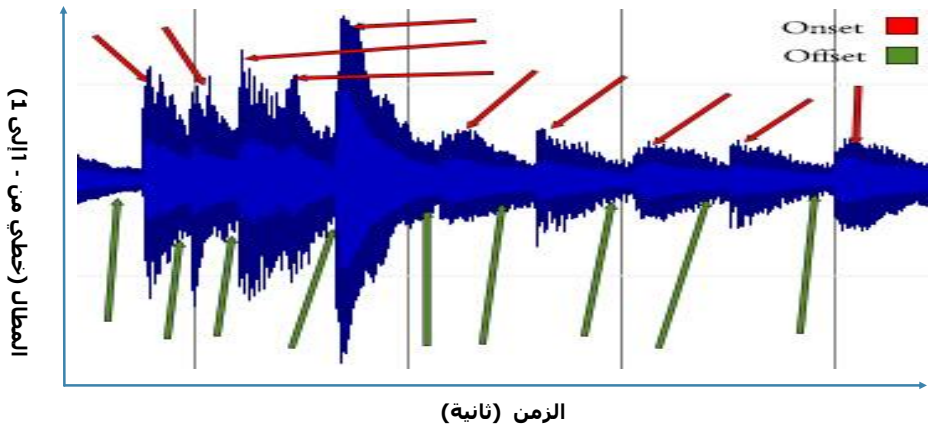
2-2 الموسيقا:

تعد الموسيقا شكلاً من أشكال الصوت المضبوط والموزون التي تهتم بتوليد وإيقاع وتوزيع الألحان، والتي تعتمد على وجود نصف بعد بين كل علامتين متتاليتين في الموسيقا الغربية لكل أوكتاف بمجمل 12 علامة في الأوكتاف الواحد في حين تعتمد الموسيقا العربية على وجود ربع بعد بين كل علامتين متتاليتين [12] كما هو موضح

في الشكل (2–1) ما يزيد عدد العلامات الموسيقية في كلّ أوكتاف من 12 حتى 24 علامة وعادة ما تكون العلامة الموسيقية إحدى هذه النغمات ذات التردد المحدد والتي تكون نشطة لفترة زمنية معينة. في حين يعرّف الكورد على أنّه ناتج تنشيط علامتين موسيقيتين مختلفتين أو أكثر ضمن الحيز الزمني نفسه. يوضح الشكل (2–2) شكل الموجة الصوتية لمقطوعة موسيقية مبينا عليها كلّ من بداية النغمة الموسيقية ونهايتها وشدة الصوت.



الشكل (2–1) يمكن أن تكون ربع العلامة الموسيقية في منتصف المسافة تماماً بين نصفي العلامة كما في (أ) أو يمكن أن تكون مزاحة قليلا عن المنتصف كما في (ب) وذلك وفقاً للآلة الموسيقية وأسلوب العزف (الباحث)



الشكل (2–2) مخطط يبين بداية ونهاية وشدة العلامات الموسيقية في جزء من مقطوعة موسيقية ولّد باستخدام برنامج Sonic Visulizer (الباحث)

3-2 واجهة الآلات الموسيقية الرقمية Musical Instrument Digital Interface:

تعرف واجهة الآلات الموسيقية الرقمية (MIDI) على أنّها معيار تقني يمثل بروتوكول اتصال وواجهة رقمية ووصلة كهربائية تصل ما بين مجموعة واسعة من الآلات الموسيقية الكهربائية وجهاز الحاسب لتشغيل وتعديل وتسجيل الموسيقا [13]. وهي ليست ملفات صوتية، بل تتكون من مجموعة من أوامر التشغيل والإطفاء بصورة أساسية ولها أحداث زمنية متنوعة أبرزها:

1- حدث بداية العلامة الموسيقية Onset event.

2- حدث انتهاء العلامة الموسيقية Offset event.

3- حدث انحناء العلامة الموسيقية Pitch bend event ويستخدم لتغيير تردد العلامة الأساسية ما يتيح لنا الحصول على ربع العلامة التي تستخدم في الموسيقا الشرقية.

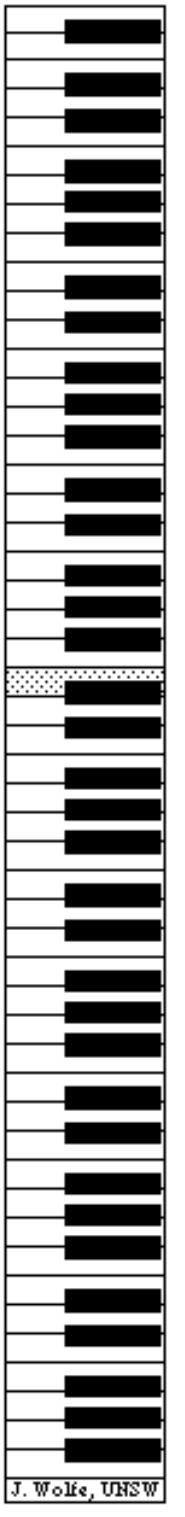
كما أنّ واجهة MIDI تسمح بتحديد قيمة شدة العلامة الموسيقية برقم مكون من سبع خانات وعادة ما تحمل ملفاتها امتداد (MID).

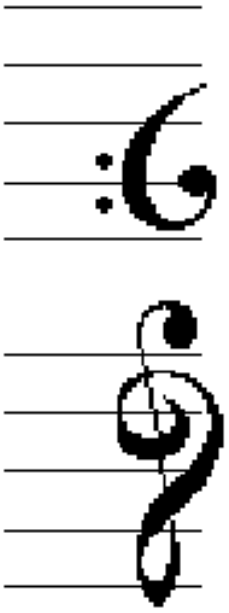
الشكل (2-3) يوضح تمثيل ملف MIDI الزمني على شكل لفات بيانو (Piano roll) (الباحث)

تدعم واجهة MIDI وجود 128 علامة مختلفة وتمثَّل كل علامة موسيقية برقم من 0 حتى 127 (بدون واحدة) كما هو موضح في الشكل (2–4)، إلا أنّنا لن نستخدم سوى المجال من 21 حتى – 108 وهو ما يوافق 89 علامة موسيقية موجودة في البيانو (غالبا ما يستخدم ما تبقى من المجال للآلات الإيقاعية)، ويمثل كل رقم من هذه الأرقام ترددا معينا كما هو موضح في الشكل (2–3).

تمثل الأرقام السابقة علامات موسيقية غربية فقط ولتمثيل العلامات الشرقية ضمن واجهة MIDI سنقوم باستخدام حدث انحناء العلامة الموسيقية والذي يمكننا من تغيير تردد العلامة الموسيقية المستهدفة ضمن مجال يتراوح بين 0 و16383 ويمثل بـ 14 خانة؛ إذ أنّ القيم الأعلى من 8192 تزيد من تردد العلامة المستهدفة بينما تقوم القيم الأدنى بتخفيض تردد العلامة، ويقع مجال تغير التردد ما بين تردد العلامة السابقة والتالية للعلامة الموسيقية المستهدفة.

MIDI number	Note name	Frequency Hz	Period ms
21	A0	27.500	36.36
22		29.135	34.32
23	B0	30.868	32.40
24	C1	32.703	30.58
25		34.648	28.86
26	D1	36.708	27.24
27		38.891	25.71
28	E1	41.203	24.27
29	F1	43.654	22.91
30		46.249	21.62
31	G1	48.999	20.41
32		51.913	19.26
33	A1	55.000	18.18
34		58.270	17.16
35	B1	61.735	16.20
36	C2	65.406	15.29
37		69.296	14.29
38	D2	73.416	13.62
39		77.782	12.86
40	E2	82.407	12.13
41	F2	87.307	11.45
42		92.499	10.81
43	G2	97.999	10.20
44		103.83	9.631
45	A2	110.00	9.091
46		116.54	8.581
47	B2	123.47	8.099
48	C3	130.81	7.645
49		138.59	7.216
50	D3	146.83	6.811
51		155.56	6.428
52	E3	164.81	6.068
53	F3	174.61	5.727
54		185.00	5.405
55	G3	196.00	5.102
56		207.65	4.816
57	A3	220.00	4.545
58		233.08	4.290
59	B3	246.94	4.050
60	**C4**	**261.63**	**3.822**
61		277.18	3.608
62	D4	293.67	3.405
63		311.13	3.214
64	E4	329.63	3.034
65	F4	349.23	2.863
66		369.99	2.703
67	G4	392.00	2.551
68		415.30	2.408
69	**A4**	**440.00**	**2.273**
70		466.16	2.145
71	B4	493.88	2.025
72	C5	523.25	1.910
73		554.37	1.804
74	D5	587.33	1.703
75		622.25	1.607
76	E5	659.26	1.517
77	F5	698.46	1.432
78		739.99	1.351
79	G5	783.99	1.276
80		830.61	1.204
81	A5	880.00	1.136
82		932.33	1.073
83	B5	987.77	1.012
84	C6	1046.5	0.9556
85		1108.7	0.9020
86	D6	1174.7	0.8513
87		1244.5	0.8034
88	E6	1318.5	0.7584
89	F6	1396.9	0.7159
90		1480.0	0.6757
91	G6	1568.0	0.6378
92		1661.2	0.6020
93	A6	1760.0	0.5682
94		1864.7	0.5363
95	B6	1975.5	0.5062
96	C7	2093.0	0.4778
97		2217.5	0.4510
98	D7	2349.3	0.4257
99		2489.0	0.4018
100	E7	2637.0	0.3792
101	F7	2793.0	0.3580
102		2960.0	0.3378
103	G7	3136.0	0.3189
104		3322.4	0.3010
105	A7	3520.0	0.2841
106		3729.3	0.2681
107	B7	3951.1	0.2531
108	C8	4186.0	0.2389





الشكل (2-4) يوضح رقم MIDI وكلا من اسم العلامة الغربي ومفتاح البيانو والتردد وطول الموجة الموافق له [14]

2-4 أطوار النغمة الموسيقية:

تمر النغمة الموسيقية بأربعة أطوار من انطلاقها وحتى نهايتها في غالبية الآلات الموسيقية. وتمثل هذه المراحل تغير مطال الإشارة الصوتية بالزمن ويمكن تلخيصها بأربعة مراحل:

1- **طور التنشيط (الهجوم) Attack:** وهو الزمن اللازم للصوت حتى يصل لمطال معين وغالبا ما يكون هذا المطال هو المطال الأعظمي. ويبدأ هذا الطور عند بداية إصدار النغمة (كلحظة النقر على مفتاح البيانو أو وتر العود مثلا). غالباً ما يكون طور التنشيط قصيراً ومحدوداً بمدّة النقر كما في البيانو والطبل والغيتار والعود، في حين يكون زمن التنشيط في الآلات النفخية كالفلوت والناي أطول من سابقاتها. ولا يكون زمن التنشيط الطويل جداً موجوداً في الآلات الموسيقية الفيزيائية؛ بل محصوراً في الآلات الموسيقية الإلكترونية والمؤثرات الصوتية فقط.

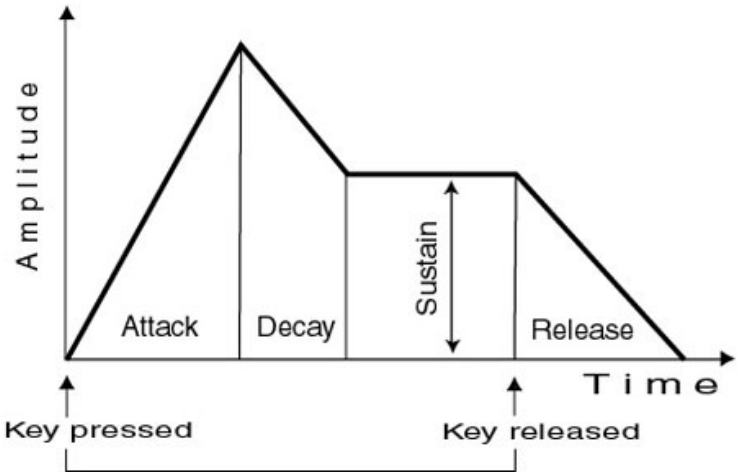
2- **طور التخامد Decay:** يبدأ هذا الطور بعد انتهاء طور التنشيط مباشرة. وهو الزمن الواقع بين طور التنشيط والامتداد وعادة ما يكون مطال الإشارة الصوتية في هذا الصوت أقل من مطالها في طور التنشيط. من ثمّ فإنّ طور الإشارة يتخامد في هذه المرحلة.

3- **طور الاستدامة Sustain:** وهو الزمن الذي يكون فيه مطال الإشارة الصوتية ذو مستوى مستقر بعد طور التخامد. ولا توجد مدّة زمنية محددة لهذا الطور؛ بل هي مرتبطة بوقت رفع اليد عن المفتاح في البيانو مثلاً أو بنهاية العلامة بصورة طبيعية.

4- **طور التحرير Release:** حالما يتم رفع اليد عن العلامة فإنّ الإشارة الصوتية تدخل في طور التحرير وفيه يحصل تخامد كبير للإشارة حتى تصل لمرحلة الصمت. بعض الآلات الموسيقية لديها زمن إطلاق طويل كالمثلث مثلا.

يمكننا تمثيل معظم الأصوات الموسيقية باستخدام الأطوار السابقة، وتختلف مدة هذه الأطوار تبعاً للآلة الموسيقية المستخدمة، كما أن بعض الآلات الموسيقية تحوي فقط طوري التنشيط والإطلاق. ويستعرض الشكل (2-5) هذه المراحل.

إنّ النظام المستخدم بحاجة لمعرفة الفروق بين هذه الأطوار على اختلافها وذلك لتحديد كلّ من بداية ونهاية النغمة الموسيقية تحديداً صحيحاً.



الشكل (2-5) يوضح الأطوار المختلفة للنغمة الموسيقية [15]

2-5 معالجة الإشارة الرقمية (DSP) Digital Signal Processing:

وهو مجال هندسي يركز على الطرائق الحسابية لتحليل ومعالجة الإشارات الرقمية المتقطعة. عادة ما تكون الإشارة الرقمية (كما في الملفات الصوتية مثلا) ناتجة عن إشارة تشابهية بحيث تُؤخذ العينات منها وفق تردد تقطيع معين. يجب على تردد التقطيع أن يكون أكبر من ضعفي تردد الإشارة التشابهية، وهو ما يدعى بشرط شانون ونايكويست كي نتمكن من تمثيلها رقميا تمثيلاً دقيقاً. في حال كان تردد التقطيع صغيرا فإنّ العديد من معلومات الإشارة التشابهية سوف تُفقد في أثناء عملية التقطيع.

2-5-1 تحويل فوريه المتقطع Discreet Fourier Transform (DFT):

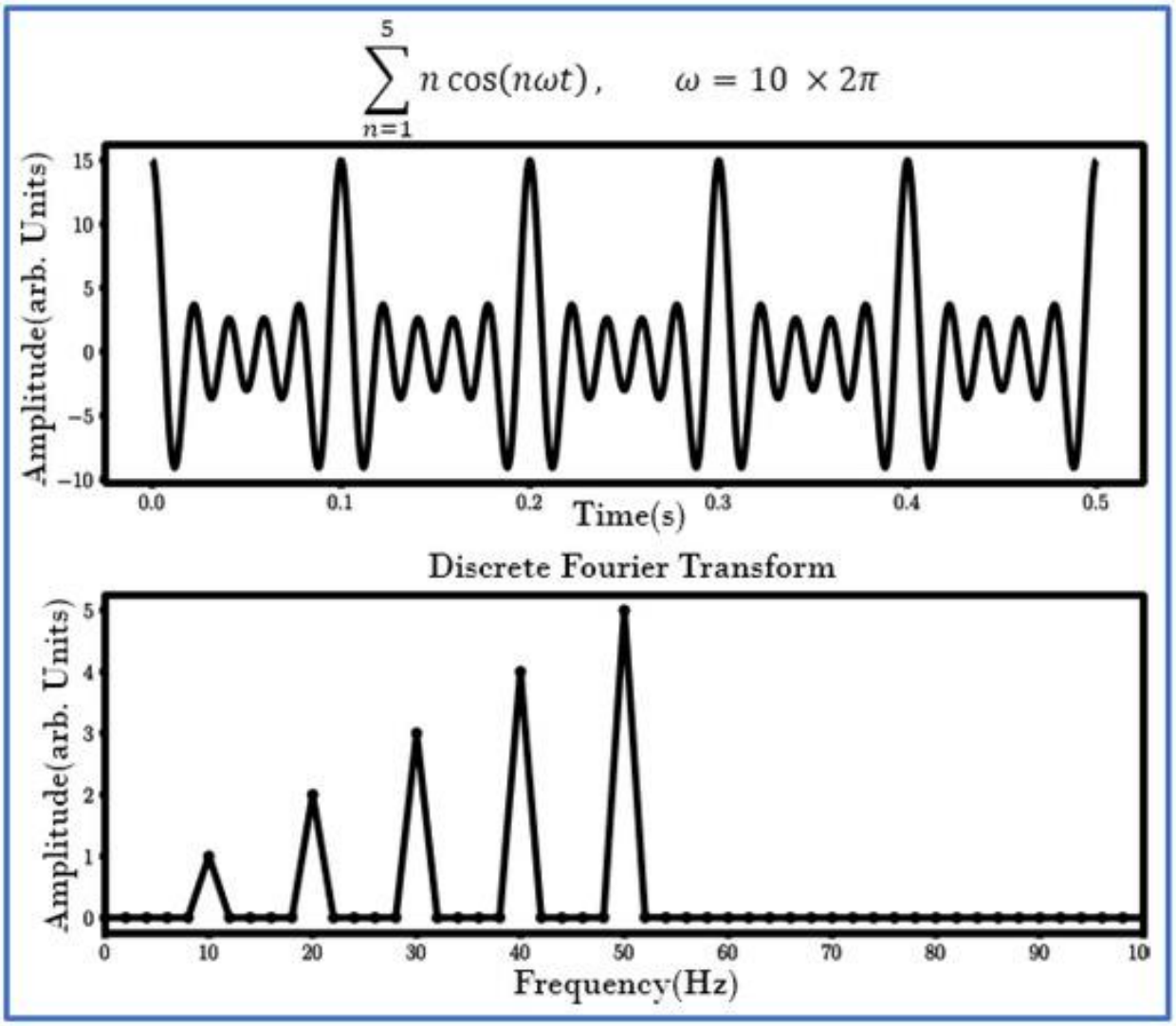
بما أنّ الحاسب لا يستطيع التعامل إلا مع الإشارات الرقمية فقد تم إيجاد تحويل فوريه المتقطع كطريقة رقمية لإيجاد تحويل فوريه. ويعد هذا التحويل أحد أهم التحويلات في مجال معالجة الإشارة المتقطعة، كما أنّ لتحويل فوريه المتقطع أهمية كبيرة في كثير من التطبيقات مثل الترشيح الخطي وتحليل الطيف وحساب الترابط.

يعطى تحويل فوريه المتقطع لإشارة دورية $x[n]$ ، مكونة من N نقطة حيث $n= 1,2, \dots , N-1$ بالعلاقة (2-1):

$$X[k] = DFT[x[n]] = \sum_{n=0}^{N-1} x[n] e^{-\frac{2\pi jkn}{N}} \qquad (2-1)$$

إذ إنّ $k= 1,2, \dots , N-1$. ومن الجدير بالذكر أنه من الشائع لحساب تحويل فوريه المتقطع استخدام تحويل فوريه السريع Fast Fourier Transform (FFT). ولحساب نتيجة التحويل يجب أن يكون حجم العينات من قوى العدد 2. ويخفض تحويل فوريه السريع التعقيد من n^2 ليصبح من رتبة $n * \log (n)$ ما يقلل من زمن المعالجة بشكل كبير.

يوضح الشكل (2-6) مثالاً عن تحويل فوريه السريع لإشارة تحوي مجموعة من المركبات الترددية المختلفة ونلحظ أنّ التحويل يجري على كامل الإشارة؛ أي إنّه يعرض المركبات الترددية في الإشارة كاملة دون تحديد زمن بداية ونهاية كل تردد.

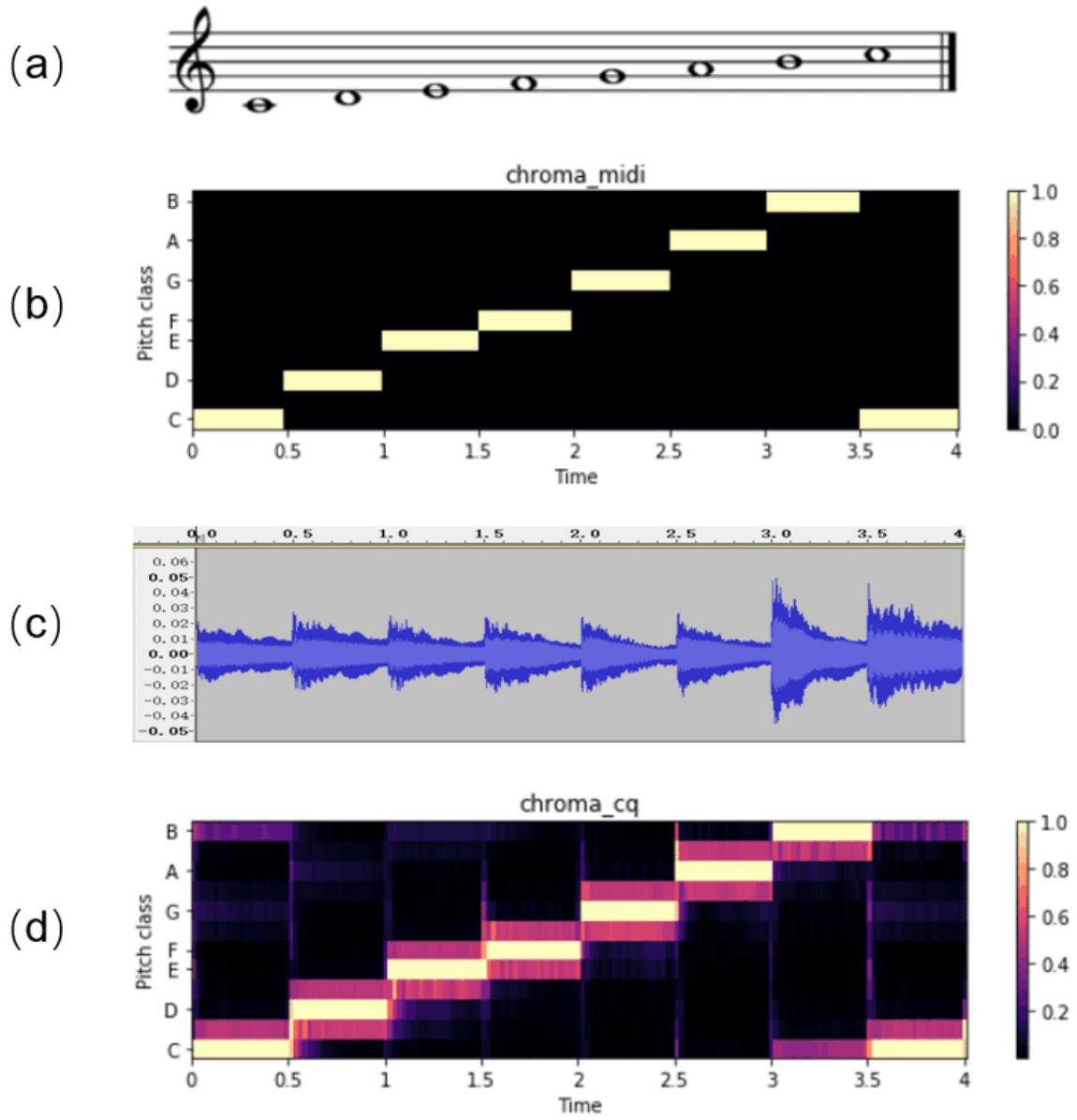


الشكل (2-6) تحويل فورييه السريع لإشارة تحوي 10، 20، 30، 40، 50 هرتز [16]

2-5-2 تحويل فورييه قصير الزمن (Short Time Fourier Transform (STFT:

رغم فاعلية تحويل فورييه المتقطع في العديد من المجالات إلا أنّه يقوم بإيجاد الطيف للإشارة كاملة. من ثمّ في حال رغبتنا بإجراء التحويل على مقطوعة موسيقية كاملة فإننا سنكون فكرة عن الترددات الموجودة في هذه المقطوعة فقط، وسيبقى المجال الزمني الذي تشغله هذه الترددات مجهولا. إنّ تحديد زمن كل نغمة موسيقية أي تردد هذه النغمة في مجال الزمن ضروري في مجال استرجاع المعلومات الموسيقية وذلك لتحديد زمن بداية ونهاية هذه النغمة.

في هذه الحالة يمكننا استعمال تحويل فورييه قصير الزمن الذي يقوم بتقسيم الإشارة كاملة إلى أجزاء صغيرة وإجراء تحويل فورييه المتقطع لكل جزء على حدة. يوضح الشكل (2–7 تشابه خرج تحليل STFT مع شكل ملف MIDI كلفات بيانو)

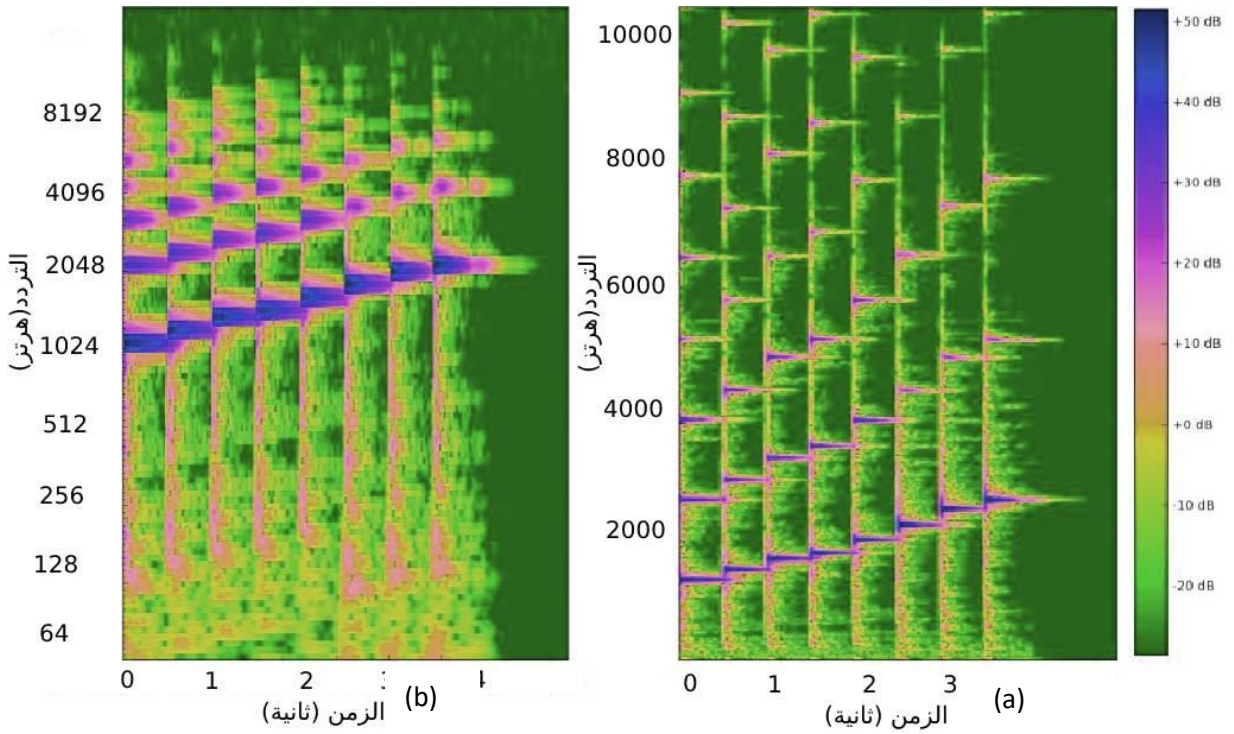


الشكل (2-7) يمثل الشكل (a) النوطة الموسيقية لسلم الميجور، الشكل (b) تمثيل بصيغة MIDI لسلم الميجور كلفات بيانو، الشكل (c) يمثل شكل الموجة الصوتية للسلم، الشكل (d) يمثل تحويل فوريه قصير الزمن [17]

2-5-3 تحويل عامل الجودة الثابت (Constant-Q-Transform (CQT:

يعد تحويل عامل الجودة الثابت أحد أبرز الطرائق لتمثيل الطيف زمنياً في الإشارات الصوتية الموسيقية، وذلك كونه مناسباً في توضيح المسافات بين التوافقيات الترددية، بخلاف تحويل فوريه قصير الزمن ذو التدرج الخطي على محور التردد، فإنّ تحويل عامل الجودة الثابت يتبع مقياسا أسّيا تكون فيه المسافات متساوية لكل

أوكتاف كما هو موضح في الشكل (2-8)؛ إذ عُرِض المخطط الطيفي لتحويل STFT على محور خطي وعرض نتائج تحويل CQT على محور أسي.



الشكل (2-8) (a) تحويل STFT ، (b) تحويل CQT لسلم الدو ميجور الموسيقي (الباحث)

ويعطى تحويل عامل الجودة الثابت بالعلاقة الرياضية (2-2) [18] :

$$X[k] = \frac{1}{N[k]} \sum_{n=0}^{N[k]-1} W[k,n]x[n]e^{\frac{-j2\pi Qn}{N[k]}}. \qquad (2-2)$$

إذ يدعى Q بعامل الجودة Quality Factor ويعطى بالعلاقة (2-3) [18]:

$$Q = \frac{1}{\delta} \qquad (3-2)$$

وتحسب δ من العلاقة (2-4):

$$\delta = (2^{\frac{1}{k}} - 1) \qquad (4-2)$$

حيث k هو عدد العلامات الموسيقية المطلوب حسابها في كل أوكتاف.

وتمثل $N[k]$ طول النافذة للعلامة الموسيقية رقم k وتعطى بالمعادلة:

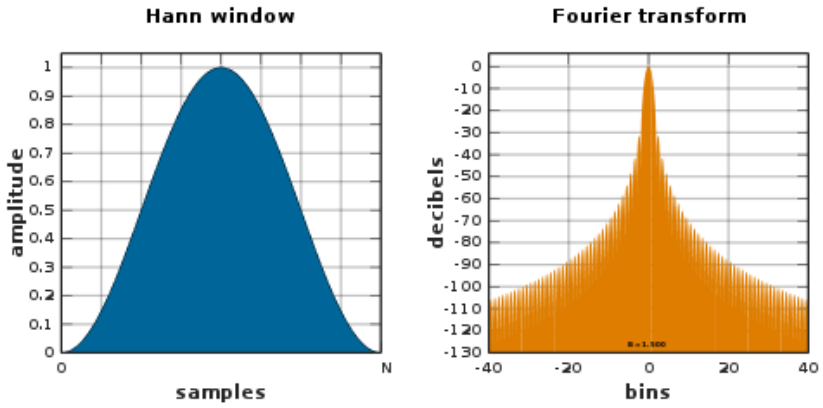
$$N[k] = \frac{f_s}{\delta f_k} = \frac{f_s}{f_k} Q. \qquad (5-2)$$

ويمثل f_s تردد التقطيع بينما يمثل f_k تردد العلامة الموسيقية المستهدفة.

تمثل $W[k,n]$ النافذة المستخدمة، وغالبا ما تستخدم نافذة Hann والتي تعطى بالمعادلة التالية:

$$W[k,n] = 0.5 - 0.5\cos\frac{2\pi n}{N[k]-1}, \qquad 0 \le n \le [k]-1 \qquad (6-2)$$

وذلك كونها تعطي دقة ترددية جيدة وتحافظ على مطال الإشارة الفعلي كما هو مذكور في [11] ويوضح الشكل (2−9) نافذة Hann والاستجابة الترددية الخاصة بها.



الشكل (2−9) نافذة Hann (يسارا) واستجابتها الترددية (يمينا).

2-6-1 اكتشاف النغمات الأحادية Mono-pitch Detection:

إنّ اكتشاف النغمات الأحادية تعد مشكلة محلولة. استُخدم تابع الترابط الذاتي بوصفه إحدى الطرائق الأولى لحل هذه المشكلة وما زال استخدامه قائما حتى الآن لتحديد التردد الأساسي ونغمة الآلة الموسيقية.

ويعرف الترابط الذاتي على أنّه ناتج ترابط الإشارة مع نسخة مؤخرة زمنياً من نفسها بوصفه تابعاً للزمن. بعبارة أخرى فإنّها تعبر عن مقدار التشابه بين الإشارة والإشارة المزاحة في كلّ نقطة من نقاط الانزياح تابعاً للزمن. إنّ مقدار التشابه في الإشارة الدورية سيكون أعظمياً عند اتفاق الإشارتين في الطور من ثمّ فإنّه يمكننا حساب تردد الإشارة الأساسية بمعرفة الزمن بين أعظم قمتين. ويعطى الترابط الذاتي للإشارة الرقمية بالعلاقة (2-7):

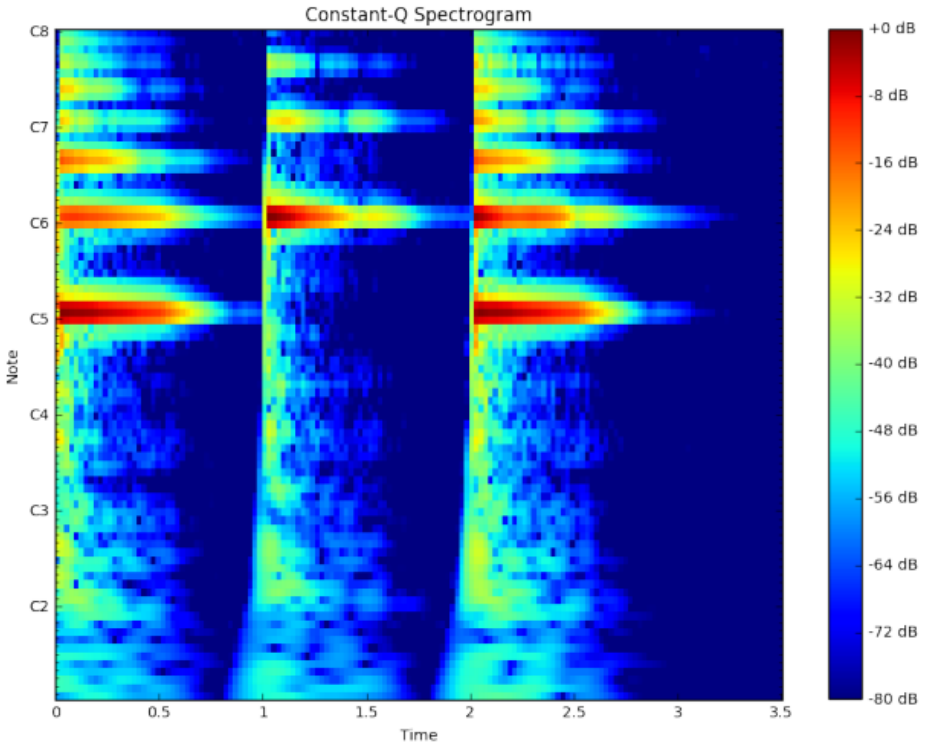
$$Rxx[j] = \sum_{n} x[n]\, x[n-j] \qquad (7-2)$$

استُخدِمت خوارزمية YIN لحساب التردد الأساسي للكلام أو للموسيقا على حد سواء[19] . وتعتمد اعتماداً أساسياً على حساب الترابط الذاتي المذكور آنفا مع بعض التعديلات لتقليل معدل الخطأ. ومن ميزات هذه الخوارزمية أنّها تملك معدل خطأ أقل بثلاثة أضعاف من الطرائق المشابهة، كما أنّه لا يوجد حد أعلى للتردد الذي يمكن حسابه من ثمّ فهي مناسبة أيضا للنغمات العالية، كما أنّها تمتاز بالبساطة وبسرعة الحساب؛ أي من الممكن استخدامها في تطبيقات الزمن الحقيقي، ومن مميزاتها أيضاً قلة المحددات اللازمة لتنفيذ الخوارزمية.

2-6-2 اكتشاف النغمات المتعددة Multi-pitch Detection:

يعد اكتشاف النغمات المتعددة أصعب التطبيقات في مجال استرجاع المعلومات الموسيقية. إنّ التحدي الأساسي يكمن عند عزف علامتين أحدهما توافقية ترددية للأخرى، من ثمّ فإنّ العلامة ذات التردد الأدنى ستعزز ترددات العلامة ذات التردد الأعلى بدلاً من ظهور قمم واضحة وجديدة في المخطط الطيفي. عند عزف نغمات عدّة سوية فإنّ بعض المركبات الترددية لهذه النغمات يمكن أن تعزز أو تلغي مركبات ترددية لعلامات أخرى كما هو

موضح في الشكل (2–10)؛ إذ تُعزَف علامة دو وحيدة في الأوكتاف الخامس (C5) ، وعلامة دو وحيدة في الأوكتاف السادس (C6) ، وعلامتي دو في كل من الأوكتافين الخامس والسادس سوية.



الشكل (2–10) نغمة دو5 منفردة يسارا، نغمة دو6 منفردة في المنتصف، نغمة دو5 ودو6 في نفس الوقت يمينا [19].

نلاحظ من هذا الشكل صعوبة التمييز بين علامة دو5 وعلامتي دو5 ودو6 المعزوفتين سوية. وتعد معرفة عدد النغمات اللازم اكتشافها في كل لحظة زمنية من المشاكل الأخرى في اكتشاف النغمات المتعددة، ففي النغمات الأحادية يوجد نغمة واحدة فقط لاكتشافها بينما يكون العدد في النغمات المتعددة مجهولا ولهذا السبب فمن غير الممكن حل هذه المشكلة باعتماد الطرائق التي تستخدم تابع الترابط الذاتي. من الصعب أيضاً كتابة شروط وقواعد يدوية لاكتشاف هذه النغمات من المخطط الطيفي في كل من تحويلي CQT وSTFT. إنّ خوارزميات تعلم الآلة والتعليم العميق تعد من الطرائق الواعدة لاكتشاف هذه الشروط والقواعد.

الفصل الثالث

تعلم الآلة والشبكات العصبونية

Machine Learning and Nurel Networks

3-1 تعلم الآلة Machin Learning:

أوجدت تقنيات معالجة الإشارة وتعلم الآلة نتائج أكثر دقة في مجال استرجاع الموسيقا وكتابة الموسيقا الآلية مقارنة بباقي التقنيات. ويعتبر تعلم الآلة فرعاً من علوم الحاسب الذي يقوم بتدريب نظام بواسطة بيانات موجودة مسبقاً ويقوم باستخدام النظام المدرب لإيجاد التوقعات والنتائج لبيانات جديدة. وتلخص مراحل تعلم الآلة بالخطوات التالية:

1- تحديد المشكلة المراد حلها.

2- تحضير البيانات

3- تدريب النظام على هذه البيانات

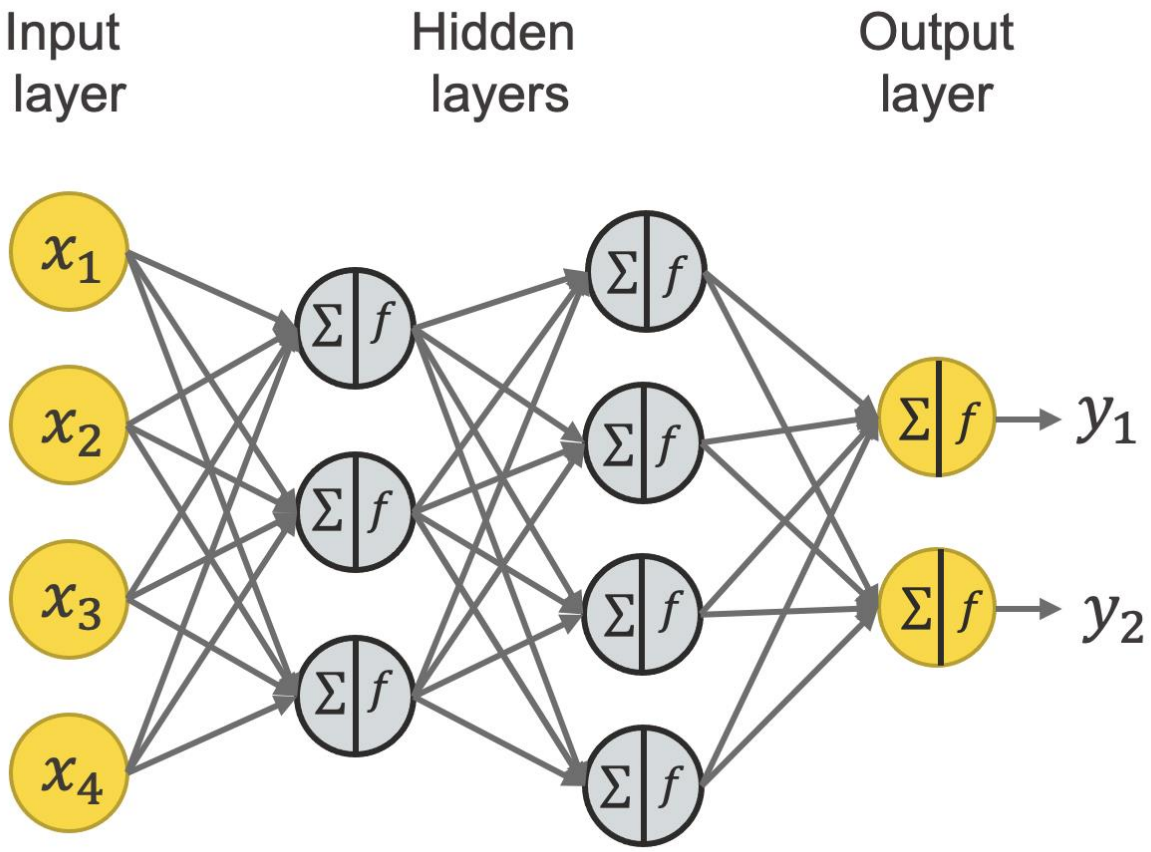
4- فحص نتائج النظام على بيانات جديدة لتعديل محددات النظام لتحقيق أداء أفضل.

سنستفيد في هذه الرسالة من تعلم الآلة لبناء نموذج تعلم آلة المشرف عليه Supervised Machine learning الذي سيستفيد من المدخلات والمخرجات على حد سواء لاكتشاف الترابط والعلاقات بينها بصورة آلية. وسيتم تكرار عملية التعليم لزيادة الدقة قدر المستطاع وسيتم اختبار النظام على بيانات لم يدرب عليها من قبل لمعرفة قدرة النظام على التعامل مع البيانات الجديدة.

3-1-1 شبكات الذكاء الصنعي:

سوف نستخدم نهج التعلم المشرف عليه في الشبكات العصبونية التي تعرف على أنّها نموذج يحاول محاكاة الدماغ بوصفه شبكةً من الخلايا العصبية والوصلات فيما بينها. سندرس بداية الشبكات العصبونية المجردة أو ما يدعى بنموذج الشبكات العصبونية كاملة الاتصال fully connected neural network model. يتكون هذا النوع من الشبكات من دخل وخرج ومجموعة أو أكثر من الطبقات المخفية. تكون كل طبقة من الشبكة العصبونية

متصلة اتصالاً كاملاً مع الطبقة التي تليها كما هو موضح في الشكل (3-1). ولكلّ من هذه الوصلات وزن خاص بها والذي يمثل محدداً متغيراً قابلاً للتعلم؛ أي أنّ قيمة هذه الوصلة تتغير مع الوقت بعملية التدريب.



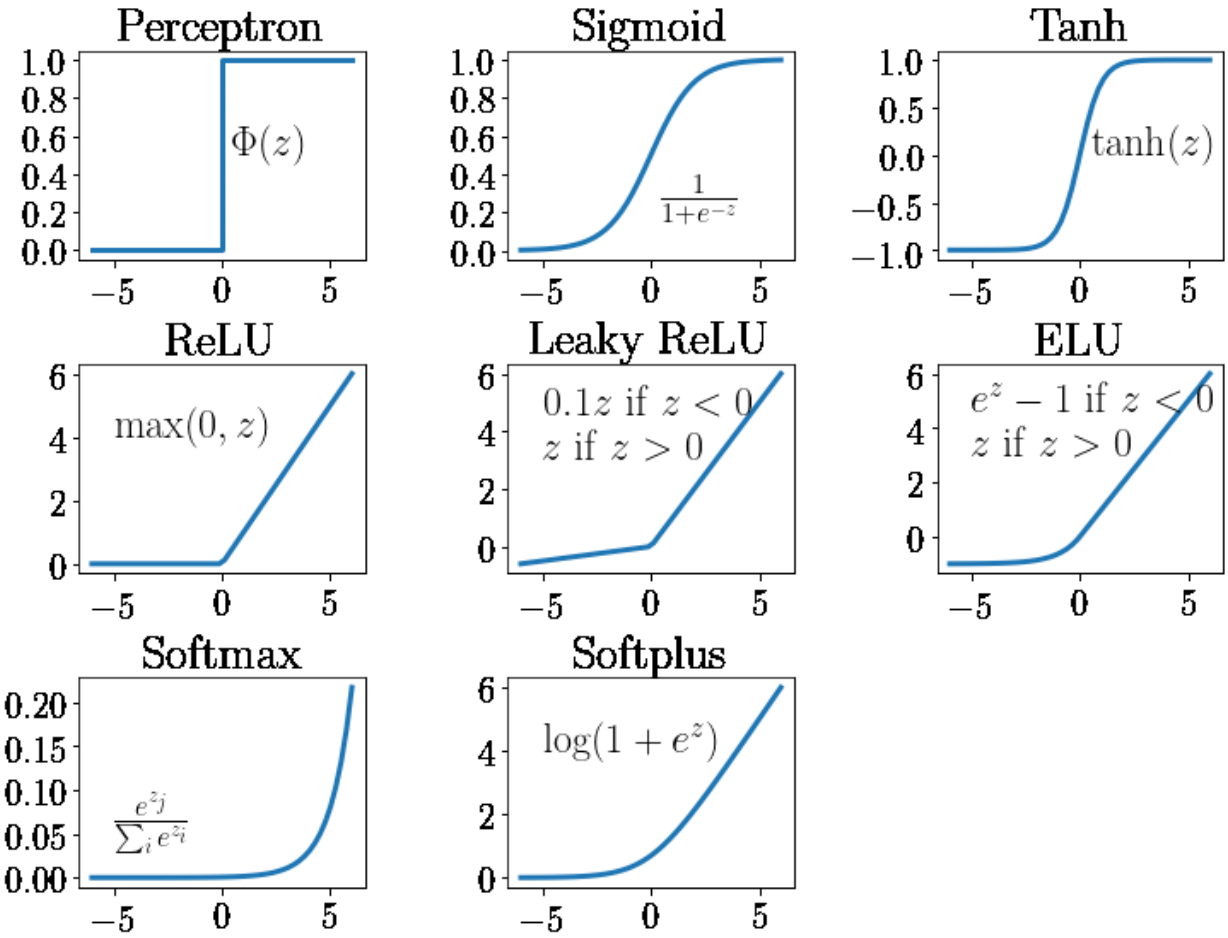
الشكل (3-1) مثال عن نموذج شبكة عصبونية كاملة الاتصال بطبقتين مخفيتين

3-1-2 الانتشار الأمامي Forward Propagation :

للوصول إلى الخرج انطلاقاً من الدخل، فإنّ الشبكات العصبونية تمر بمرحلة الانتشار الأمامي؛ إذ نقوم بضرب كل دخل مع الوزن الذي يشاركه مع كل عصبون من الطبقة المخفية التالية. كل عصبون في الطبقة التالية سيصبح عنده مجموع موزون لعصبونات الطبقة السابقة كلها. يُطبّق تابع تنشيط Activation Function على هذه المجاميع الموزونة (العصبونات) مثل تابع Sigmoid مثلاً. تتكرر هذه العملية حتى نصل لطبقة الخرج.

3-1-3 توابع التنشيط Activation Functions:

تطبق توابع التنشيط بعد ضرب قيم عصبونات الدخل والأوزان لكل طبقة كاملة الاتصال. وتقوم هذه التوابع بضغط قيم المجاميع الموزونة كلّها في مصفوفة الخرج لتصبح ضمن مجال محدد. ويوضح الشكل (3-2) أشهر توابع التنشيط. سيتم استعمال توابع تنشيط ReLU وSigmoid في النظام المصمم.



الشكل (3-2) توابع التنشيط الأكثر شيوعا [20]

3-1-4 توقع النتائج:

في مشاكل التصنيف البسيطة Simple Classification Problems، فإنّ الشبكة العصبونية ستحاول معرفة ما إذا كان الدخل ينتمي إلى صف معين Class وذلك بالاعتماد على المميزات المستخرجة بواسطة CNN.

عادة ما يكون خرج الشبكة العصبونية ذو قيمة تتراوح بين 0 و1 ويمكننا تقريب الخرج إلى 0 أو 1 وفق عتبة معينة. إنّ المشاكل البسيطة غالباً ما تنتمي إلى صف واحد فقط فإمّا أن يكون نشطا أو لا يكون (مثاله هل تحوي الصورة على إنسان؟ وجوابها نعم أم لا أي 0 أو 1)، ولكن في حالة تعدد الصفوف Multiple classes فعادة ما يُستخدم تابع تنشيط Softmax في الخرج الذي يضمن أن يكون مجموع قيم عصبونات الخرج جميعها مساوياً للواحد من ثمّ فإنّ الخرج ذو القيمة الأكبر غالباً ما يكون الصف الصحيح.

بالنسبة لمشكلتنا فإنّنا نحاول معرفة العلامات الموسيقية النشطة في كل إطار ضمن المخطط الطيفي لتحليل CQT؛ أي إنّه من الممكن أن يكون هناك أكثر من علامة موسيقية نشطة في الوقت نفسه. وهذا ما يدعى في الشبكات العصبونية بالتصنيف المتعدد النصوص Multi-label classification. في هذا النوع من المشاكل لا يمكننا استعمال تابع تنشيط Softmax لأنّنا لا نحاول اختيار العلامة الأكثر وضوحاً فقط. بدلاً من ذلك فإنّنا نحاول وضع عتبة معينة، وفي حال تجاوز قيمة عصبون الخرج المرتبط بعلامة موسيقية معينة لهذه العتبة فعندها يمكننا عدّ العلامة نشطة. تكون هذه العتبة ثابتة للعلامات جميعها وتضبط الشبكة العصبونية الأوزان بمقارنة الدخل مع الخرج الصحيح الذي تُزوّد الشبكة العصبونية به على شكل مصفوفة ثنائية البعد تمثل محوري الزمن والعلامات الموسيقية (تردد كل علامة).

3-1-5 الانتشار الخلفي Backward Propagation:

تعتمد هذه الطريقة على المقارنة بين الخرج الذي قامت الشبكة العصبونية بحسابه والخرج الصحيح المقدم للشبكة العصبونية. في حال عدم تطابق النتائج تقوم الشبكة العصبونية بحساب الفرق في كل عصبون من عصبونات الخرج وتقوم بحساب الخطأ (error) وذلك باستخدام تابع الخسارة (Loss Function). يتبع هذه المرحلة مرحلة الانتشار الخلفي التي تقوم بحساب قيمة الخطأ في كل عصبون من عصبونات الطبقات المخفية. وبعدها تُحدّث قيمة الأوزان وذلك عن طريق حساب الأوزان الجديدة لكلّ عصبون وتبديل القيم القديمة بالقيم الجديدة. يشترط

في الانتشار الخلفي أن تكون دوال التنشيط التي تستعملها العصبونات قابلة للاشتقاق. ذلك لأنّه في مرحلة تحديث الأوزان تستعمل الدالة المشتقة لدالة التنشيط في حساب القيم الجديدة [21].

3-2 الشبكات العصبونية الالتفافية (CNN) Convolutional Neural Networks:

سنقوم باستخدام الشبكات العصبونية الالتفافية بوصفها عنصراً أساسياً في بناء النموذج المستخدم في هذه الرسالة. وذلك كونها أظهرت نتائج جيدة في مهمات تصنيف الصور. يمكننا اعتبار المخطط الطيفي لتحويل CQT كصورة رمادية وذلك بسبب التمثيل الثنائي للمخطط على محوري الزمن والتردد.

تملك الشبكات العصبونية الالتفافية بنية مشابهة للشبكات العصبونية كاملة الاتصال آنفة الذكر، فكلاهما يملك عصبونات وأوزان وقيم انحياز Biases. يستقبل كل عصبون دخلاً ويجري عليه عمليات خطية وقد يتلوه بصورة اختيارية تابع تنشيط معين غالباً ما يكون ReLU في هذا النوع من الشبكات.

إنّ الشبكات العصبونية كاملة الاتصال العادية لا تملك قدرة التعرف على مميزات البيانات ذات الهيكلية المعينة كالبيانات ثنائية البعد ومثالها الصور الرمادية أو المخططات الطيفية. على فرض أنّ دخل الشبكة العصبونية كان إطاراً من المخطط الطيفي ممثل بمصفوفة ثنائية أبعادها 360×6، فإنّ عدد مميزات الخرج سيكون جداء البعدين أي 3160 مميز. وعلى فرض أنّنا استخدمنا 200 عصبون في الطبقة التالية فإن عدد الأوزان سيكون 200×3160 = 432000 وزنا. وهو عدد كبير من الأوزان الذي غالباً ما سيؤدي لمشكلة زيادة الملائمة Overfitting ولذلك يجب علينا الاستفادة من بنية الشبكة ثنائية البعد. تقوم الشبكات العصبونية الالتفافية بالاستفادة من بنية الدخل ثنائي البعد بطريقة عملية. على عكس الشبكات العصبونية العادية فإنّ بنية الشبكات العصبونية الالتفافية تكون ثلاثية البعد. وهي العرض والطول والعمق (width, height, depth)؛ إذ يمثل كلّ من العرض والطول أبعاد الصورة، في حين يمثل العمق عدد القنوات اللونية وهي ثلاث قنوات في الصورة الملونة وقناة واحدة في الصورة ذات التدرج الرمادي.

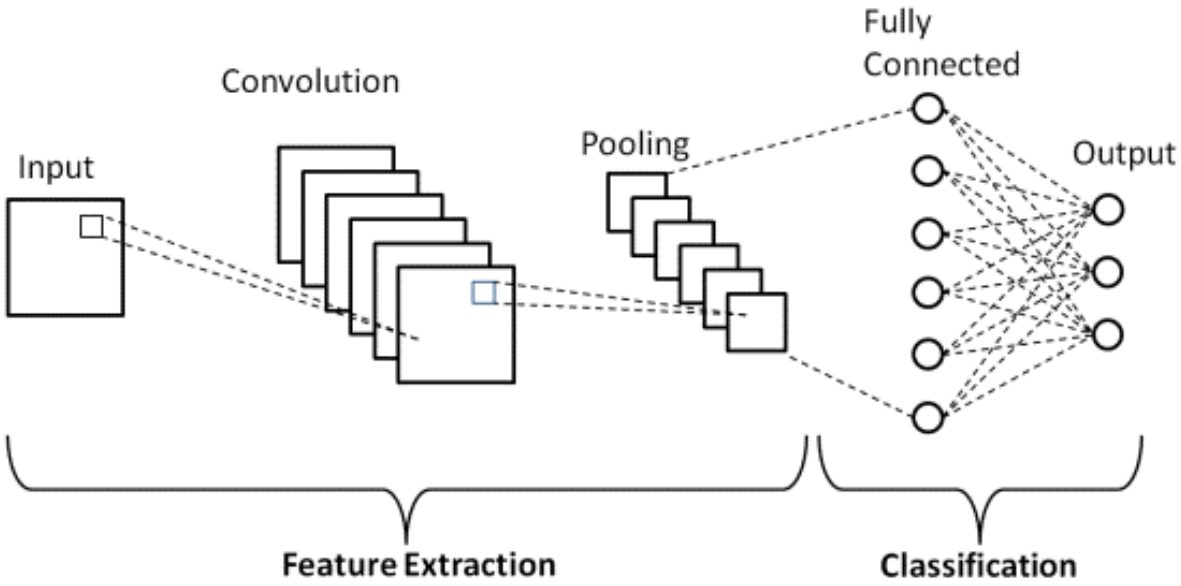
فعلى فرض استخدامنا لصورة بحجم 32×32×3 بيكسل التي تمثل العرض والطول والعمق (عدد القنوات اللونية) على التتالي، فسوف تتصل العصبونات في الشبكة العصبونية الالتفافية بجزء محدد من الطبقة السابقة فقط بدلاً من الاتصال بالعصبونات كلها كما هو الحال في الشبكات العصبونية العادية، كما أنّها ستقوم بعملية انتخاب للقيم (Pooling) ما سيخفض حجم الطبقات التالية المخفية. وسيكون حجم الخرج مصفوفة ثلاثية أبعادها 1×1×10 مثلاً؛ أي إنّ العمق سيكون واحد وذلك لتمثيل صفوف الخرج على شكل مصفوفة أحادية مرتبة وفق بعد العمق.

يوجد ثلاثة أنواع من الطبقات أو العمليات الموجودة في الشبكات الالتفافية النموذجية:

1- **الطبقة الالتفافية:** تقوم هذه الطبقة بتطبيق عدد ثابت من المرشحات في الطبقة الواحدة التي عادة ما يزداد عددها مع كل طبقة تالية بشكل التفافي على الخرج (أي تطبيق تابع رياضي على عناصر سوية عدةّ لتوليد عنصر ذو قيمة جديدة). يقوم كل مرشح بحساب ناتج ضرب الأوزان الخاصة به مع المناطق الصغيرة المطبقة على الدخل. ولهذه العملية أربع محددات أساسية وهي حجم نواة المرشح (Kernal size) وعدد المرشحات (Filters number) وحجم خطوة المرشح (Stride) والحشو بقيمة صفرية (Zero padding). يمثل حجم النواة أبعاد المرشح المطبقة على الصورة، في حين يمثل عدد المرشحات كمية المرشحات المطبقة على الصورة علما أنّ للمرشحات كلها في الطبقة الواحدة الحجم نفسه، في حين يمثل حجم خطوة المرشح عدد خطوات الانزياح للمرشح عند تطبيقه على الصورة. نستفيد من عملية الحشو بقيم صفرية لجعل كافة صور الدخل ذات حجم واحد في حال كانت أبعاده مختلفة؛ إذ يتم ملء البيكسلات الناقصة بقيمة صفرية لسد الفجوة في الحجم بين الصور المختلفة. بعد تطبيق عدد من المرشحات على الصورة يطبّق تابع تنشيط ReLU وهو التابع الأكثر شيوعاً في الشبكات العصبونية الالتفافية، ولا يغير تابع التنشيط من حجم البيانات؛ بل يغير من قيمها.

2- **طبقة الانتخاب Pooling layer:** تقوم بعملية تخفيض للتقطيع Downsampling على طول الأبعاد المكانية (الطول والعرض) بأخذ القيم المتوسطة (Average pooling) أو ذات القيمة الأكبر (MaxPooling) على فرض أن أبعاد الطبقة كانت [32x32x12] (بدون وحدة) فإنّ تطبيق عملية انتخاب أعظمي بحجم 2x2 على خرج الطبقة سيصبح حجمها [16x16x12] ونلحظ أنّ العرض والطول قد انخفض حجمهما إلى النصف دون التأثير على عدد المرشحات والذي بقي 12 مرشحا. يوضح الشكل (3-3) بنية الشبكة.

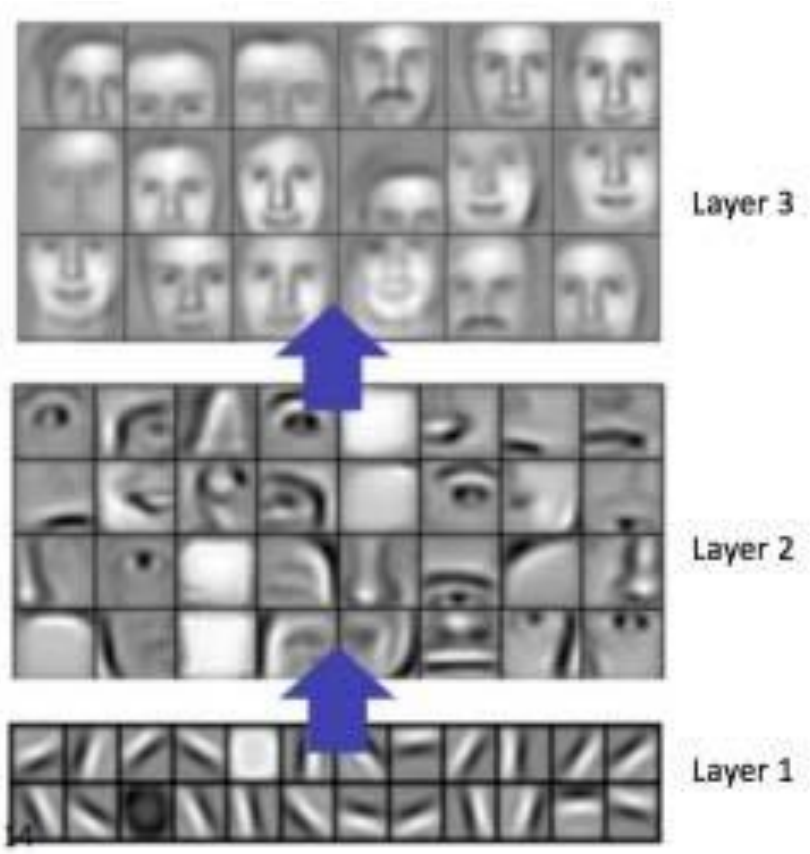
3- **طبقة من العصبونات كاملة الاتصال:** موجودة بعد الطبقات الالتفافية وطبقات الانتخاب في حال وجودها.



الشكل (3-3) بنية الشبكة العصبونية الالتفافية [22]

يمكن لشبكات CNN اكتشاف المميزات سواء في محوري الزمن أو التردد للمخططات الطيفية للمقطوعات الصوتية كما هو مقترح في [23]، إذ يعامل المخطط الطيفي معاملة الصورة الرمادية. يوضح الشكل (3-4) مثالاً عن

المميزات المكتشفة في الشبكة العصبونية الالتفافية، ونلحظ أنّ النموذج يكتشف في الطبقات الأولى مميزات أولية كالخطوط والمنحنيات، ويزداد تعقيد هذه المميزات كلما ازددنا عمقا في الطبقات؛ إذ يبدأ باكتشاف العيون والأنف والحواجب والفم، وفي الطبقات الأخيرة يكتشف النموذج معظم مميزات الوجه.



الشكل (3-4) المميزات المكتشفة من قبل الشبكة العصبونية في نظام لتصنيف الوجوه [24]

3-3 التعلم العميق Deep Learning:

تم اقتراح مصطلح التعلم العميق عام 2010 الذي يعني تكديس طبقات الشبكات العصبونية في بنية هرمية بهدف بناء خريطة للمميزات. لم يكن هناك داع لاستخدام ما يزيد عن طبقتين في الشبكات العصبونية التقليدية؛ إذ

يمكننا تمثيل أي تابع غير خطي بطبقتين فقط. يختلف الأمر بالنسبة لتمييز الصور وذلك لأنّ البنية الهرمية وزيادة الطبقات تساعد الشبكة العصبونية على اكتشاف المزيد من المميزات.

كان استخدام التعلم العميق في السنوات السابقة حكراً على الحواسب ذات قدرة المعالجة العالية وذلك بسبب الكم الهائل من العمليات الحسابية المطلوبة وكمية العصبونات والأوزان الكبيرة المستخدمة. مكنت وحدات معالجة البيانيات Graphic processing units GPUs من تنفيذ هذه العمليات على الحواسب الشخصية وذلك بسبب قدرتها الكبيرة على إجراء العمليات الحسابية على المصفوفات بصورة تفرعية بخلاف وحدات المعالجة المركزية Central Processing Units CPUs والتي تستغرق وقتاً طويلاً. من المناسب استعمال تابع تنشيط ReLU في شبكات التعلم العميق ولا سيّما شبكات CNN إذ أنّ احتمال بقاء الأوزان في القيم العظمى أو الصغرى دون تغير قيمها أقل من توابع Tanh وSigmoid، كما أنّ تابع التنشيط من نوع ReLU يزيد من سرعة تدريب الشبكة العصبونية ستة أضعاف كما هو مذكور في [25]. وتبين الدراسة ذاتها أن تابع تنشيط ReLU أظهر نتائج أفضل في مجال الصوتيات.

الفصل الرابع

جمع بيانات التدريب والمعالجة المسبقة للعينات الموسيقية

Dataset Collecting and Musical Samples Preprocessing

4-1 مقدمة:

يتعامل الحاسب مع الملفات الرقمية فقط، كما أنّ بنية الشبكة العصبونية بحاجة إلى دخل مخصص مناسب لها وإلا فلن تستطيع معالجة البيانات. من ثمّ فإنّنا بحاجة لعملية معالجة مسبقة للبيانات لتغيير شكل الملفات الصوتية إلى شكل تستطيع الشبكات العصبونية فهمه. تشمل هذه العمليات تحويل البيانات الصوتية وملفات MIDI إلى مصفوفات أحادية أو ثنائية تكون دخلاً للشبكة العصبونية. كما أنّه يجب إجراء عملية تقييس Normalization لضبط القيم جميعها ضمن مجال معين كي تؤثر على تغير قيم الأوزان تأثيراً منطقياً. يتطلب الأمر أيضا إجراء بعض التحليلات الرياضية وذلك لاستخلاص بعض المميزات وتسهيل اكتشافها من قبل الشبكة العصبونية [27]. إنّ كل ما سبق يسهم في تحسين دقة النتائج وفي تسريع عملية التدريب وسنستعرض في هذا الفصل الخطوات التي قمنا بها لمعالجة البيانات قبل تمريرها للشبكة العصبونية.

4-2 قاعدة بيانات التدريب Training Dataset:

إنّ أنواع الشبكات العصبونية التي تستخدم نهج التعلم المشرف عليه لتغيير الأوزان بحاجة لكلّ من الدخل والخرج الصحيح المتوقع لهذا الدخل كي تضبط الاوزان بصورة صحيحة. وذلك عن طريق الانتشار الأمامي والخلفي آنفي الذكر. لذلك فنحن بحاجة لقاعدة بيانات تدريب تحوي كلاً من الملفات الصوتية التي نستطيع إجراء التحليلات الرياضية لتحويلها إلى مخطط طيفي، وملفات المعلومات المرتبطة بهذه الملفات الصوتية التي توضح زمن بداية ونهاية وتردد كل علامة موسيقية.

تبين لنا بعد البحث المطول أنّ عدد قواعد بيانات التدريب للموسيقا العربية خصوصاً والشرقية عموماً محدود. رغم توفر بعض قواعد البيانات للموسيقا التركية (لها بنية مشابهة للموسيقا العربية) إلا أنّها تفتقر لتعدد النغمات من ثمّ فهي غير صالحة لمشكلتنا كوننا نحاول اكتشاف عدة نغمات في نفس الوقت، كما أنّ النغمات الشرقية غالباً ما تكون متذبذبة ضمن ربع العلامة الموسيقية (يعزى ذلك إلى التغير السريع والمفاجئ في حدث pitch

bend الموضح في الفقرة (2-3))؛ أي أنّ تردد العلامة يتغير تغيراً كبيراً وواضحاً بالنسبة للمستمع وهو ما قد يؤثر سلبا بصورة كبيرة على أداء النموذج المصمم. إنّ قواعد البيانات السابقة الموجودة لا تصلح لحل مشكلتنا، كما أنّ إنشاء قاعدة بيانات تدريبية جديدة من البداية يحتاج إلى وقت طويل والعديد من الموارد لذلك قمنا بابتكار حلّ لهذه المشكلة ضمن الموارد المتاحة.

4-2-1 قاعدة بيانات (MAPS) MIDI Aligned Piano Sounds:

وهي قاعدة بيانات غربية للبيانو تحوي على ملفات MIDI وملفات WAV المقابلة لها كما أنّها تحوي على ملفات نصية لكلّ من بداية العلامة ونهايتها ورقم MIDI الخاص بها. وتحوي عدداً كبيراً من الملفات الصوتية سواء لنغمات موسيقية معزولة ومعزوفة بسرعات مختلفة، أو لكوردات موسيقية متنوعة، أو لمقطوعات موسيقية كاملة ويبلغ حجمها 15 GB. اعتمدت كثير من الأبحاث السابقة على قاعدة البيانات هذه للموسيقا الغربية ولكن تكمن المشكلة الأساسية بأّنه لا توجد علامات شرقية لتدريب الشبكة العصبونية عليها.

4-2-2 إنشاء قاعدة بيانات للموسيقا الشرقية:

إنّ عملية إنشاء قاعدة بيانات من الصفر تستغرق زمناً ومجهوداً كبيرين وتحتاج إلى فريق متخصص من العازفين والخبراء ولا يمكن إنجازها من قبل فرد واحد. كما أنّها بحاجة إلى تجهيزات موسيقية تقنية مخصصة. لحل هذه المشكلة قمنا بكتابة برنامج باستخدام لغة C# باستعمال مكتبة DryWetMIDI [27] لمعالجة ملفات MIDI

وتتكون هذه العملية من عدة مراحل هي:

1- تعديل زمن امتداد النغمة الموسيقية.

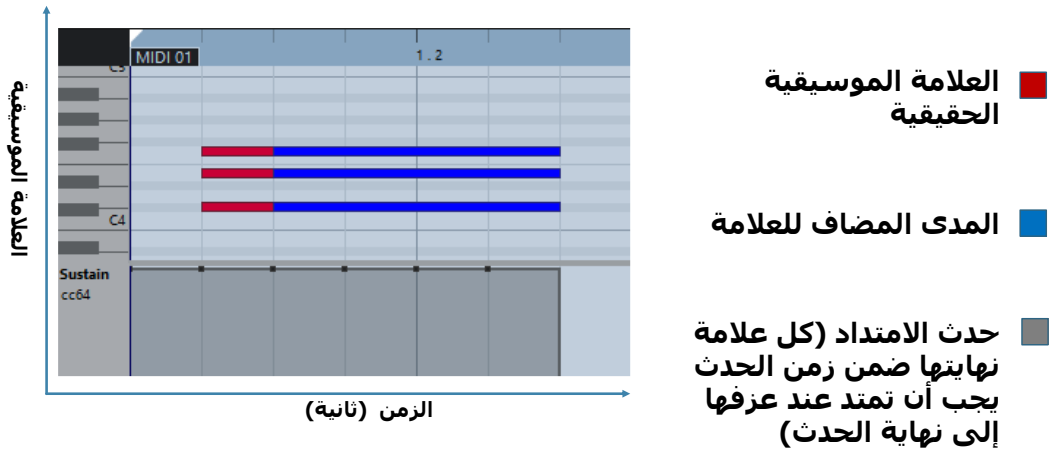
2- تحويل الموسيقا الغربية إلى موسيقا شرقية.

3- تحويل المقطوعات الموسيقية المولدة لصيغة تفهمها الشبكة العصبونية.

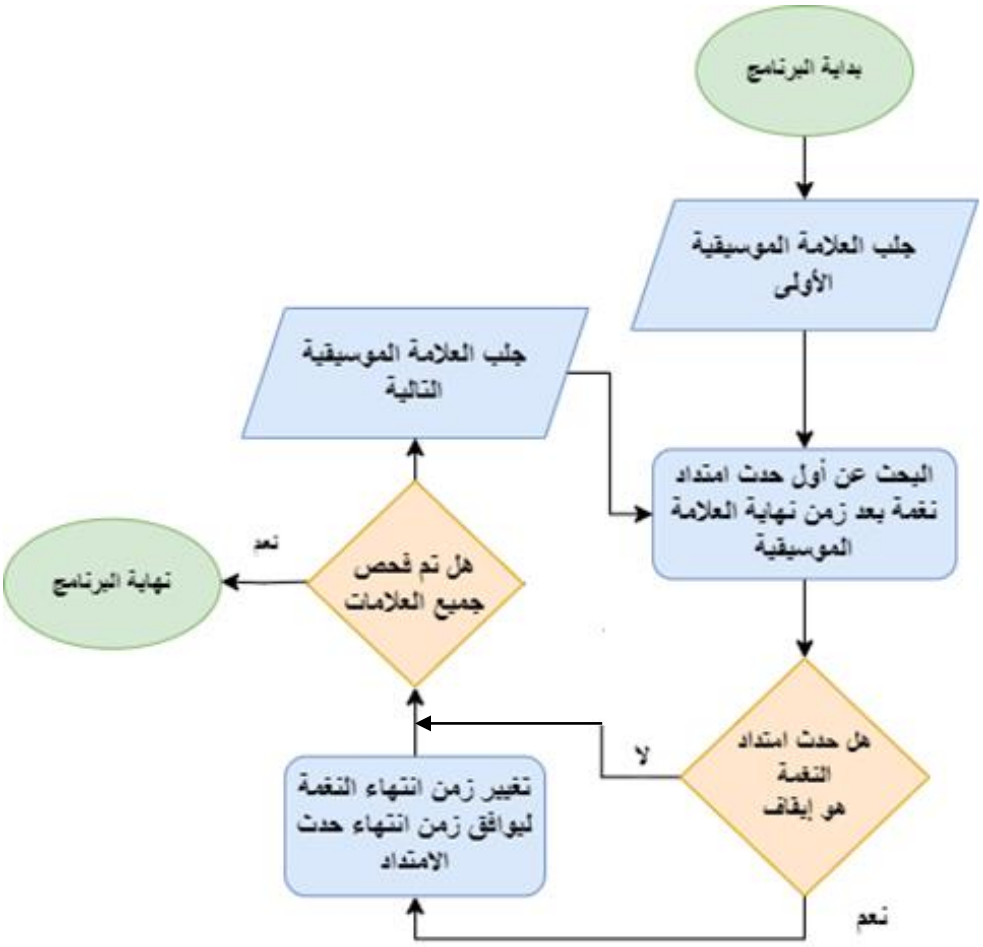
4-2-2-1 تعديل زمن امتداد النغمة الموسيقية:

من الأحداث الموجودة في صيغة MIDI ما يدعى بحدث Sustain pedal أو دواسة امتداد النغمة وهو تمثيل لزر موجود في آلات البيانو الموسيقية ويستخدم كي يزيد من زمن امتداد العلامة الموسيقية وله الوظيفة نفسها في ملفات MIDI. ويتفعل هذا الحدث في حال كانت نهاية نغمة موسيقية تقع ضمن مجاله؛ إذ يزيد طول العلامة الموسيقية حتى ينتهي الحدث كما هو موضح في الشكل (4-1).

إنّ الخطوة الأولى من خطوات المعالجة هو إلغاء حدث sustain pedal واستبداله بعلامات أطول زمنياً وفق الخوارزمية الموضحة في الشكل (4-2).



الشكل (4-1) حدث امتداد النغمة الموسيقية (الباحث)



الشكل (4-2) خوارزمية تعديل طول العلامة المستعملة (الباحث)

4-2-2-2 تحويل العلامات الغربية إلى علامات شرقية:

كما ذكرنا سابقاً فإنّ الفرق الرئيسي بين الموسيقا الشرقية والغربية هو وجود ربع العلامة الموسيقية في الموسيقا الشرقية؛ أي إنّ عدد العلامات الموسيقية يزداد من 12 علامة حتى 24 علامة. إنّ الموسيقا العربية تتبع سلالم موسيقية مختلفة تسمى بالمقامات الموسيقية وجميعها مشتقة من المقامات الموسيقية الأساسية السبعة وقد يضيف البعض إليها مقام الكورد فتصبح ثمانية وتكون المسافة بين العلامات فيها إمّا بعد أو نصف بعد أو ثلاثة أرباع البعد وهذه المقامات هي:

1- **مقام البيات:** ودرجة ركوزه الأساسية هي نغمة ري. وعلاماته كالآتي (ري – مي نصف بيمول – فا – صول – لا).

2- **مقام الصبا:** ودرجة ركوزه الأساسية هي نغمة ري. وعلاماته كالآتي (ري – مي نصف بيمول – فا – صول بيمول – لا – سي بيمول – دو – ري بيمول).

3- **مقام الراست:** درجة ركوزه الأساسية هي نغمة دو. وعلاماته كالآتي: (دو – ري – مي نصف بيمول – فا – صول – لا – سي نصف بيمول – دو).

4- **مقام السيكاه:** وعلامة ارتكازه هي المي نصف بي مول وعلاماته كالآتي: (مي نصف بيمول – فا – صول – لا بيمول – سي – دو – ري – مي نصف بيمول).

5- **مقام الحجاز:** وعلامة ارتكازه هي الري. وعلاماته كالآتي (ري – مي بيمول – صول بيمول – صول – لا – سي بيمول – دو – ري).

6- **مقام العجم:** وهو يوافق سلم الدو ميجور الغربي وعلامة ارتكازه هي الدو. علاماته كالآتي (دو – ري – مي – فا – صول – لا – سي – دو)

7- **مقام النهاوند:** ويوافق سلم الدو ماينور الغربي وعلامة ارتكازه هي الدو. علاماته كالآتي (دو – ري – مي بيمول – فا – صول – لا بيمول – سي – دو).

8- **مقام الكورد:** وعلامة ارتكازه هي الري. تعطى علاماته كالآتي (دو – ري – مي – فا – صول – لا سي بيمول –دو).

إنّ مقامات (البيات والصبا والراست والسيكاه) جميعها تحوي أرباع أبعاد بين بعض العلامات أي أنها ذات طابع شرقي من ثم فهي بحاجة لتعديل النغمة. في حين مقامات العجم والنهاوند والحجاز والكورد لا تحمل أرباع أبعاد من ثم لها علامات الموسيقا الغربية نفسها أي إنّها ليست بحاجة لتعديل.

كما ذكرنا سابقاً فإنّ أرقام MIDI تمثل علامات غربية فقط، ولحل هذه المشكلة قمنا باختيار 60 مقطوعة موسيقية بصيغة **MIDI** وتجاهلنا الملفات الصوتية التي تحوي العلامات الموسيقية المنفردة والكوردات كونها لا تحاكي الظروف الموجودة على أرض الواقع. قُسِّمت هذه العلامات الموسيقية لثلاثة مجموعات كلّ مجموعة تتكون من 20 مقطوعة موسيقية. حاولنا أن يكون مجموع أزمنة المقطوعات الموسيقية متساوياً قدر الإمكان:

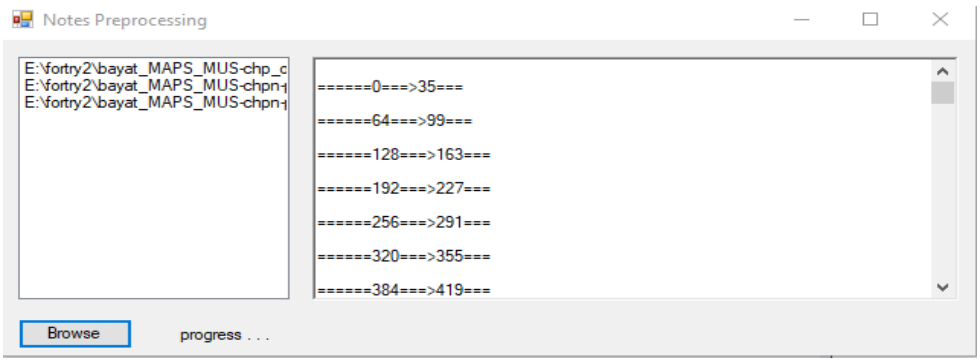
1- **المجموعة الأولى:** قمنا بإضافة حدث إنحناء علامة موسيقية للمقطوعات الموسيقية كلّها في المجموعة بقيمة **6144** لكل علامة مي موسيقية وهو ما يوافق مقامي **البيات والصبا**.

2- **المجموعة الثانية:** قمنا بإضافة حدث إنحناء علامة موسيقية لكافة المقطوعات الموسيقية في المجموعة بقيمة **6144** لكلّ من علامات مي وسي وهو ما يوافق مقامي **السيكاه والراست**.

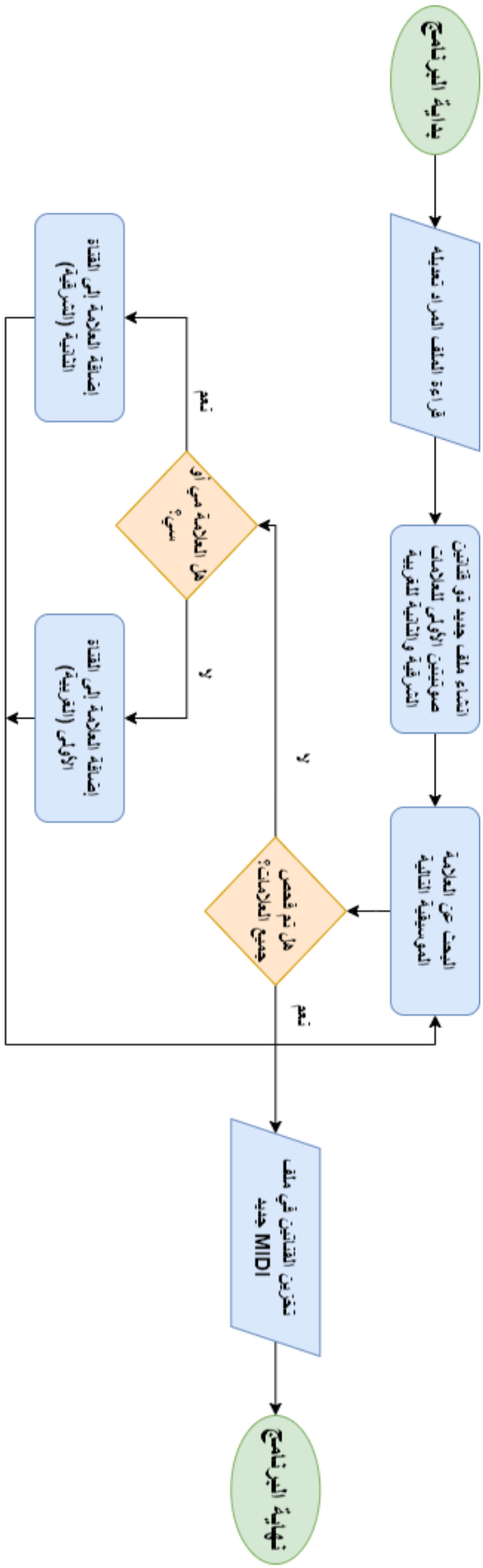
3- **المجموعة الثالثة:** لم يُدخل حدث انحناء العلامة الموسيقية في هذه المجموعة لأنّ ما تبقى من مقامات شرقية تحوي العلامات نفسها في الموسيقا الغربية وهي **الكورد والنهاوند والعجم والحجاز**.

هذا النوع من المعالجة يسمح لنا بالحصول على مقطوعات موسيقية تحتوي المقامات الموسيقية الشرقية الأساسية وعدداً كبيراً من المقامات المشتقة منها.

قمنا بتنفيذ العمليات السابقة بواسطة برنامج تمت كتابته بلغة **C#** وواجهته مبينة في الشكل (4-3) باستعمال مكتبة DryWetMIDI، ويوضح الشكل (4-4) الخوارزمية المستعملة لتنفيذ هذه العملية.



الشكل (4-3) واجهة برنامج تعديل قاعدة بيانات MIDI (الباحث)



الشكل (4-4) خوارزمية تعديل المقام الموسيقي (الباحث)

4-2-2-3 تحويل مقاطع MIDI إلى ملفات صوتية:

بعد التعديلات السابقة فإنّ ملفات WAV المرفقة مع قاعدة بيانات MAPS لن تكون صالحة لتدريب الشبكة العصبونية وذلك بسبب تغير زمن العلامات الموسيقية الناتج عن التعديل وتخفيض بعض العلامات بمقدار ربع علامة موسيقية.

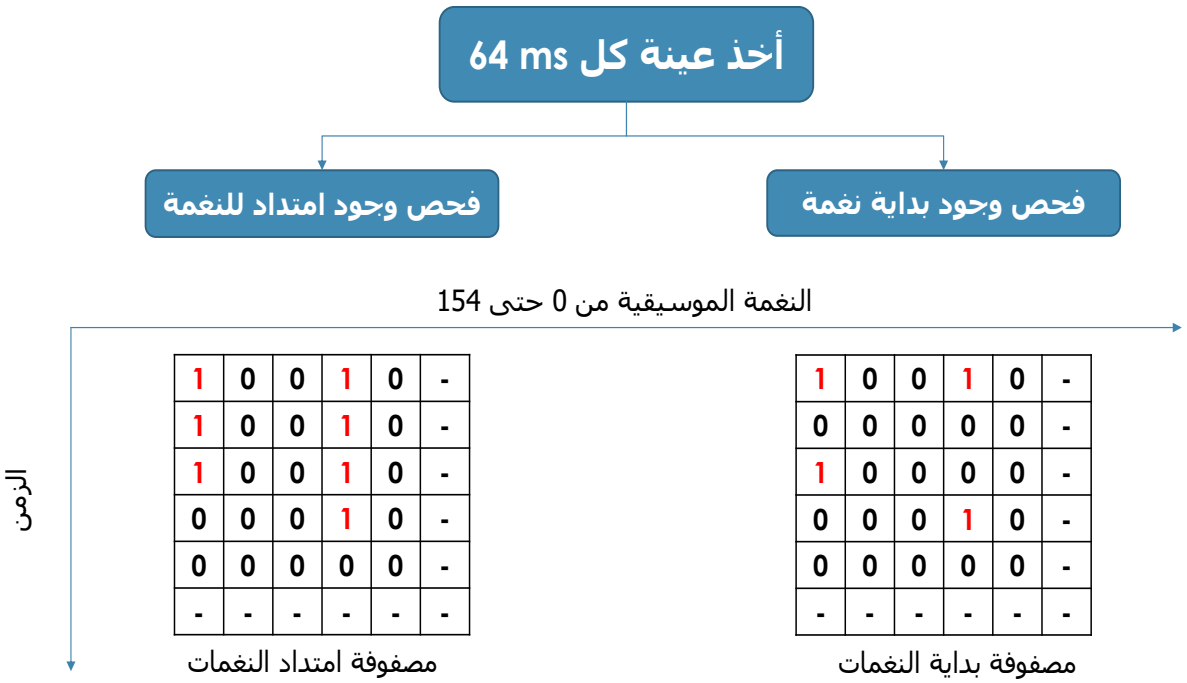
لحلّ هذه المشكلة قمنا باستخدام برنامج Raper الخاص بتعديل ملفات MIDI وقمنا باستخدام مركِّب Upright Piano synthesizer لتحويل ملف MIDI إلى ملف WAV بقناة واحدة mono channel وتردد تقطيع 44100 Hz و 24 bit لتمثيل تعديل رمز النبضة Pulse Code Modulation (PCM) بطول يوافق طول ملف MIDI الأصلي؛ أي أنّ الملفين متزامنان. بالنسبة لعينات التقييم فقد قمنا باستخدام بيانو مركِّب مختلف Dpiano synthesizer يعطي صوت بيانو ذو طابع صوتي مختلف عن المركب المستعمل لتوليد عينات التدريب.

4-2-3 تحويل صيغة MIDI لمصفوفة ثنائية:

تحتاج الشبكة العصبونية إلى صيغة خاصة لفهم المدخلات وتحليلها لذلك وجب علينا تحليل ملفات MIDI لقاعدة بيانات التدريب وتحويلها إلى مصفوفة ثنائية تستطيع الشبكة العصبونية فهمها.

يعتمد النظام المصمم على اكتشاف بداية النغمة على حده واستخدام البيانات الناتجة عنه للمساعدة في اكتشاف امتداد النغمة الموسيقية أيضا. من ثمّ فإنّنا بحاجة لمصفوفتين لتدريب الشبكة العصبونية؛ إذ تمثل المصفوفة الأولى مصفوفة بداية النغمات الموسيقية بينما تمثل المصفوفة الثانية مصفوفة امتداد النغمة الموسيقية. يتم عمل مسح لكامل ملف MIDI وذلك بتقسيمه لخطوات كل خطوة طولها 64 ms. يُفحص وجود بداية نغمة موسيقية بهذا الإطار الزمني وفي حال وجود حدث بداية نغمة يتم وضع رقم "1" في الحقل الموافق من الجدول وفي حال لم يوجد حدث يتم وضع رقم "0" إذ يمثل المحور الأفقي العلامات الموسيقية بطول 155 في حين يمثل المحور العمودي محور الزمن (كل تدريجة 64 ms) وهو ذو طول متغير حسب طول المقطوعة الموسيقية. على غرار اكتشاف النغمة

الموسيقية، فإنّ اكتشاف امتداد النغمة يتم بطريقة مشابهة ويتميز امتداد النغمة في الجدول بتتالي السلاسل ذات القيمة "1" على محور الزمن. يوضح الشكل (4-5) عملية المسح وكيفية تعبئة الجداول.



1	0	0	1	0	-
0	0	0	0	0	-
1	0	0	0	0	-
0	0	0	1	0	-
0	0	0	0	0	-
-	-	-	-	-	-

1	0	0	1	0	-
1	0	0	1	0	-
1	0	0	1	0	-
0	0	0	1	0	-
0	0	0	0	0	-
-	-	-	-	-	-

الشكل (4-5) آلية تحويل ملف MIDI إلى مصفوفتي بداية وامتداد النغمة، وتوضيح لشكل الخرج (الباحث)

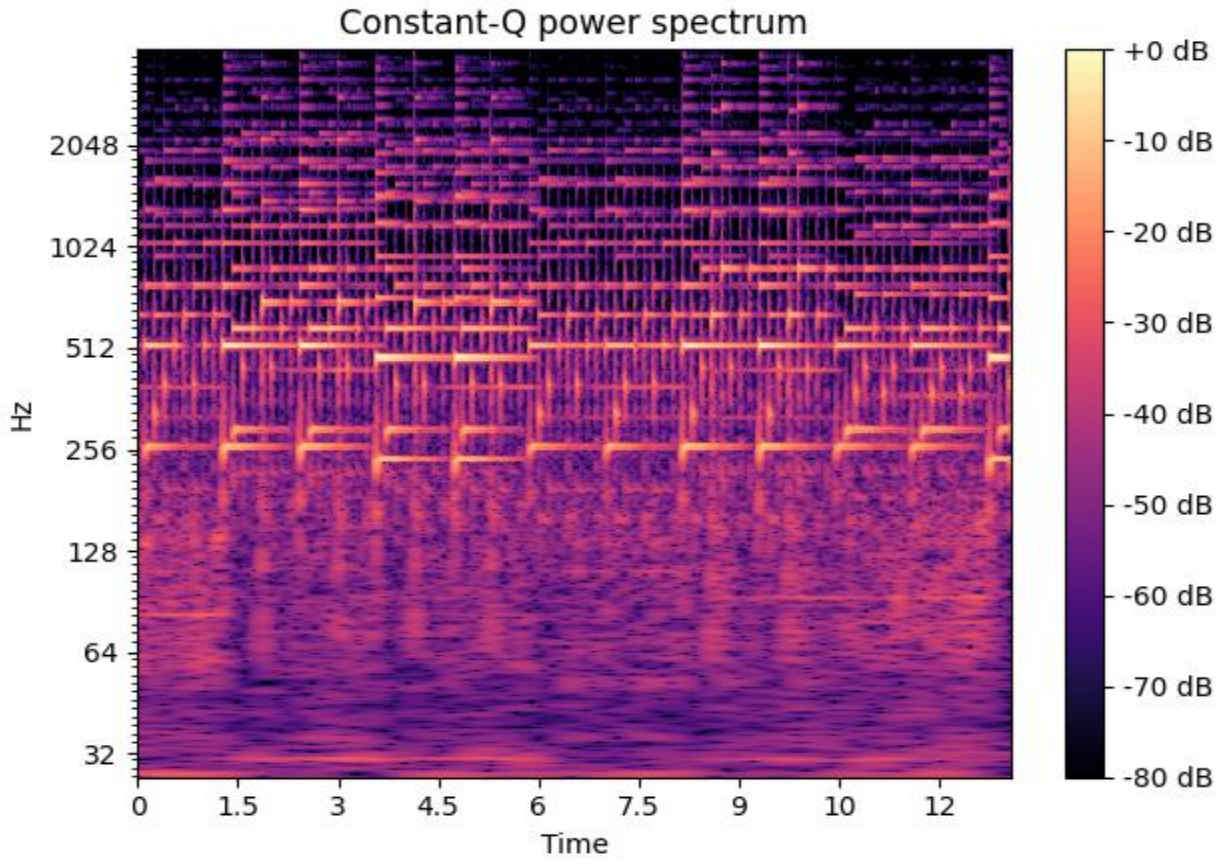
تحفظ المصفوفتان السابقتان بصيغة CSV ويتم إدخالهما للشبكة العصبونية. في هذه الحالة يصبح لدينا لكل ملفMIDI ثلاث ملفات مرفقة وهي ملف الـ WAV وملفا CSV؛ تقدم هذه الملفات جميعها دخلاً للشبكة العصبونية. يمثل ملف WAV الدخل الذي نود توقع حالته واكتشاف علاماته الموسيقية في حين تَعدُّ الشبكة العصبونية ملفي CSV على أنّهما الخرج المتوقع والصحيح لهذا الدخل في أثناء عملية التدريب، من ثمّ فهناك خرجان للشبكة العصبونية.

4-3 معالجة ملفات WAV:

تحوي صيغة WAV الأساسية معلومات الإشارة في المجال الزمني وهي تسلسل من عينات PCM ولتحويلها لشكل تستطيع الشبكات العصبونية فيه استخلاص المميزات منه يجب علينا تحويلها إلى المجال الترددي، كما يلزم علينا معالجة البيانات رياضيا وتقييسها بما يتناسب مع الشبكة العصبونية وفق الخطوات التالية.

4-3-1 تنفيذ تحليل CQT على الملفات الصوتية:

يقوم تحليل CQT بتمثيل التغيرات الترددية للمقطوعة الموسيقية على محور الزمن كما هو موضح في الشكل (4-6) المولد باستخدام لغة بايثون Python علماً بأنّ التحليل مطبق باستخدام مكتبة Librosa [28]. يُحسب تحويل CQT وفق المحددات الموضحة في الجدول (4-1).



الشكل (4-6) الطيف الناتج عن تحويل CQT لـ 12 ثانية من المقطوعة MAPS_MUS-bach_846_AkPnBcht.wav من قاعدة بيانات MAPS (الباحث)

جدول (4-1) المحددات المستخدمة في تحليل CQT (الباحث)

البارامتر	القيمة	الواحدة
تردد التقطيع	16000	Hz
عدد العلامات في الأوكتاف	48	-
عدد العلامات الكلي	344	-
طول القفزة	256	عينة
التردد الأدنى	32	Hz
التردد الأعلى	4186	Hz
النافذة المستعملة	hann	-

نقوم أولاً بإجراء عملية Downsampling لكل ملف صوتي ليصبح 16000 Hz بواسطة مكتبة Librosa إذ إنّ أعلى تردد في البيانو هو 4186 Hz من ثمّ فإنّ تردد التقطيع يحقق شرط شانون ونايكويست بأن يكون تردد التقطيع أعلى من ضعف تردد العلامة وذلك لضمان عدم ضياع المعلومات، كما أنّه أصغر تردد مدعوم من قبل كروت الصوت ويحقق هذا الشرط وتم اختياره لتسريع عملية التدريب وتخفيف الحمل الحسابي على المعالج ذو قدرة المعالجة المحدودة. إن طول قفزة 256 يعطينا زمنا مقداره 256/16000 ما يساوي 16ms؛ أي تُجمّع كل 4 قفزات متتالية سوية لتصبح بطول 64ms وهو مساو لزمن أخذ العينات لبداية وامتداد العلامة الموسيقية في ملف MIDI. وفقا لـ [29] فإنّ العلامة الموسيقية المتوقعة يجب أن تكون ضمن مجال زمني بقيمة $\mp 100\, ms$ من العلامة الأصلية كي تُعدّ صحيحة، لذلك تم اختيار أقل قيمة من مضاعفات 2^n تحقق هذا الشرط. تكون أبعاد خرج التحليل مصفوفة ثنائية بحجم $344 \times \left(\frac{\text{طول المقطوعة بالميلي ثانية}}{64}\right)$.

4-3-2 معالجة الطيف الناتج عن تحليل CQT:

تستطيع الشبكة العصبونية استخلاص بعض المميزات من ناتج التحليل السابق إلا أنّها ستواجه صعوبة في اكتشاف المميزات المعقدة، كما أنّ المجال الواسع للقيم سيؤدي إلى إبطاء عملية التعلم وسنصبح بحاجة لمعدل تعلم عال لضبط الأوزان وهو ما قد يزيد من معدل الخطأ. لحل هذه المشاكل فإننا بحاجة لإجراء العمليات الرياضية التالية:

1- **تحويل قيم مصفوفة خرج CQT إلى ديسيبل:** يساعد التحويل إلى ديسيبل في تكبير المجال بين القيم العالية والمنخفضة من ثم يساعد الشبكة العصبونية على اكتشاف الفروق بين النغمات النشطة والنغمات الخاملة ويجري التحويل باستخدام مكتبة Librosa باستخدام التعليمة البرمجية التالية:

```
CQTdbtest  =librosa.amplitude_to_db(C2, ref=np.max)
```

ويمثل np.max القيمة العليا في مصفوفة خرج CQT تُحسب القيم بالديسبل بالمعادلة التالية:

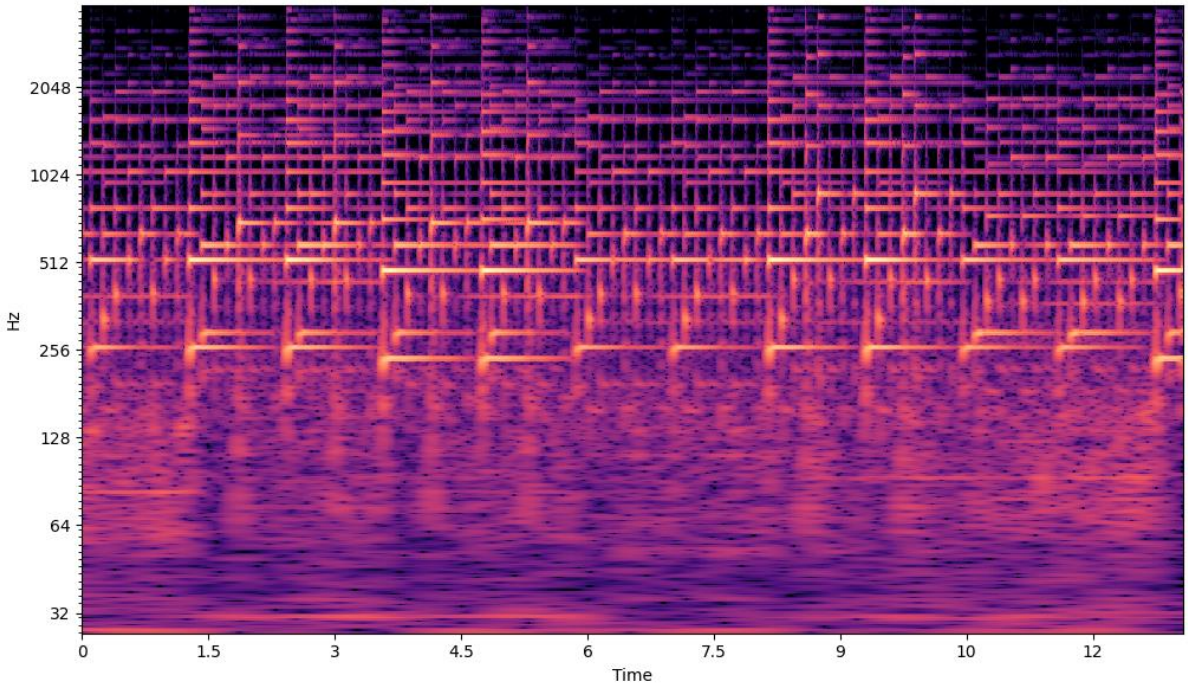
If scalar, the amplitude `abs(S)` is scaled relative to `ref`: `20 * log10(S / ref)`. Zeros in the output correspond to positions where `S == ref`.

إذ تمثل S القيمة المراد تحويلها. في حال كانت S تساوي ref فإن الناتج سيكون صفر ديسبل.

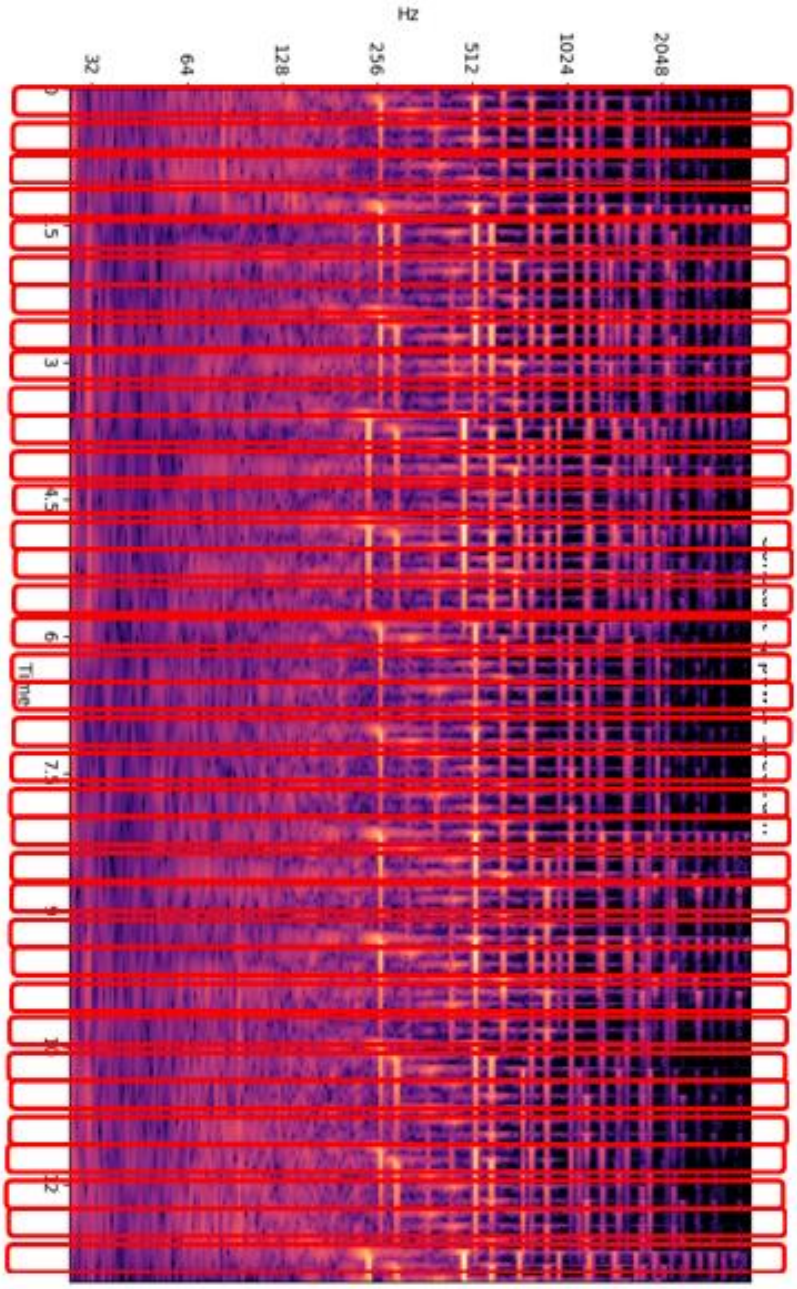
2- **عملية تغيير مجال القيم Rescale:** كما هو وارد في [30] فإن تغيير قيم المجال لتصبح من 0 إلى واحد يؤدي لتسريع عملية التعلم. نُفّذت هذه العملية بلغة البرمجة بايثون وذلك بجمع 80 لكافة عناصر المصفوفة ثم تقسيمهم على 80 مرة أخرى.

```
dfCQTdBtest = dfCQTdBtest + 80
dfCQTdBtest = dfCQTdBtest / 80
```

3- **حساب منقول المصفوفة:** تقوم الشبكات العصبونية الالتفافية بتطبيق المرشحات ابتداء من الزاوية اليسرى العليا للمصفوفة وتقوم بمسح كل سطر وبعد الانتهاء من مسحه تنتقل نحو السطر التالي. كما ذكرنا فإن خطوة تحليل CQT ذات طول 16ms تُجمّع كل أربع خطوات معاً لتكوين صورة أبعادها 4×344 بيكسل إذ يمثل رقم 4 محور الزمن في حين يمثل الرقم 344 محور التردد. لكي نستطيع تقسيم الصورة فإنّنا بحاجة إلى حساب منقولها ليصبح محور الزمن هو المحور الشاقولي (وهو ما يوافق محاور مصفوفتي بداية النغمات وامتداد النغمات) كما هو موضح في الشكلين (4-7) و (4-8).



الشكل (4–7) الطيف الناتج عن تحويل CQT قبل حساب منقول المصفوفة (الباحث)



الشكل (4-8) مثال عن الطيف الناتج عن تحويل CQT بعد حساب منقول المصفوفة وتجميع كل أربع إطارات معا (الباحث)

4- **تغيير أبعاد المصفوفة كي تناسب دخل الشبكة العصبونية:** كما ذكرنا سابقاً فإنّ لشبكات CNN ثلاثة أبعاد وهي الطول والعرض والعمق. يمثل تحويل CQT صورة ذات تدرج رمادي؛ أي في هذه الحالة تكون قيمة العمق واحد. لتمرير المخطط الطيفي الناتج يجب تقسيمه إلى صور متعددة ولتحقيق ذلك نضيف بعدا

رابعا يمثل الزمن (تجميع كل أربع إطارات متتالية معاً بوصفها صورةً مستقلةً وإدخالها للشبكة العصبونية بترتيب زمني معين) ويختلف طوله من مقطوعة موسيقية لأخرى ليصبح شكل دخل الشبكة العصبونية كالتالي (**T,344,4,1**) وتمثل كلا من عدد الصور وعدد المستويات الترددية وعدد العينات الزمنية والعمق على التتالي من اليسار لليمين.

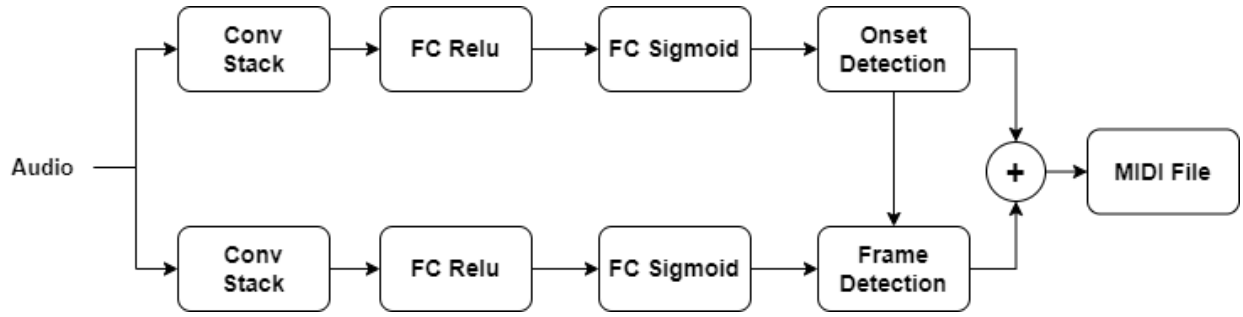
الفصل الخامس

تصميم النموذج والنتائج

Model Design and Results

5-1 بناء النموذج:

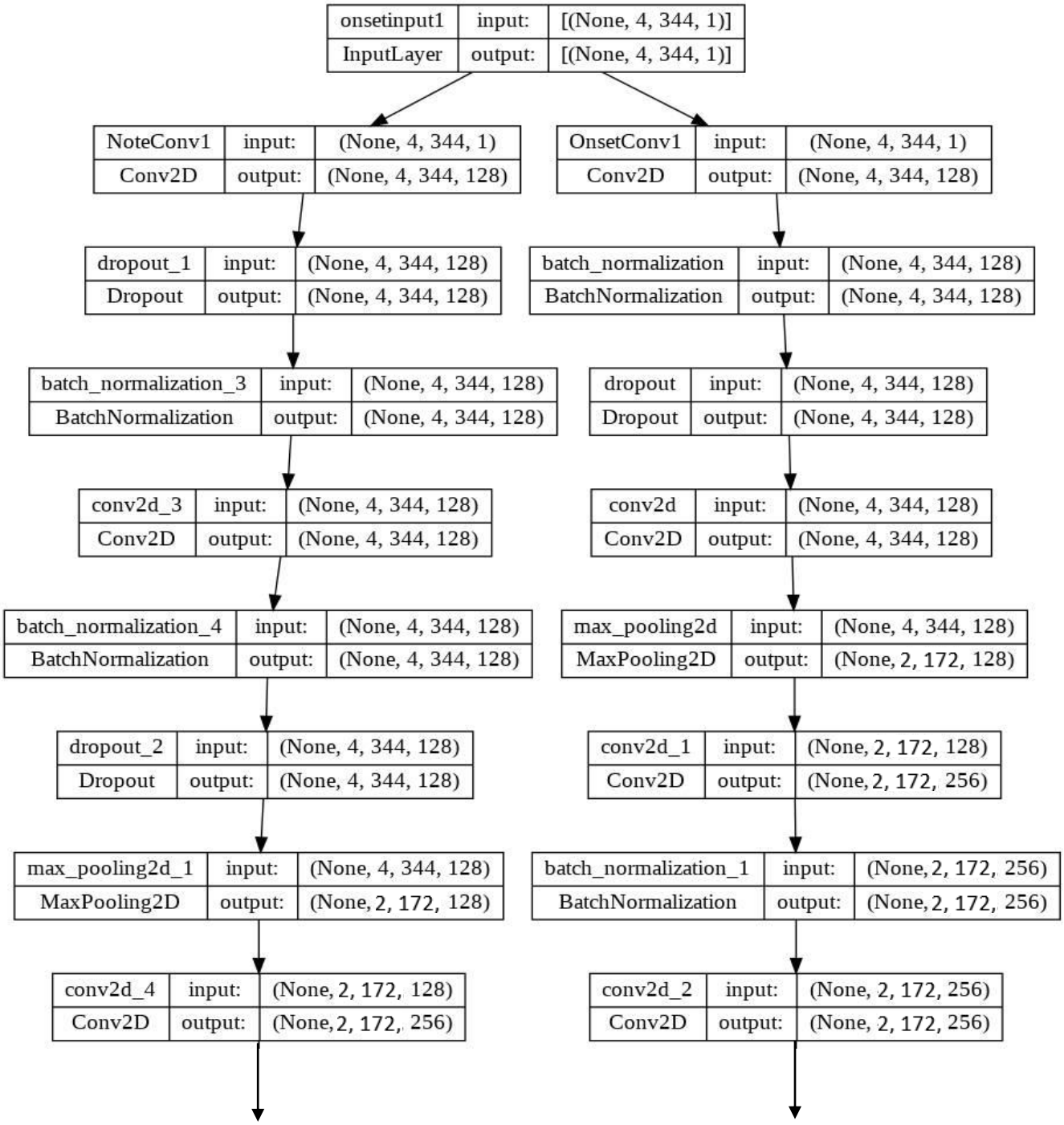
تم بناء النموذج باستخدام لغة البرمجة بايثون Python الإصدار 3.9 نسخة 64 bit ولبناء الشبكة العصبونية استُخدِمت كل من مكتبتي Tensorflow [31] الإصدار 2.7.0 و Keras Aplication programming Interface [32] الإصدار 2.7.0 التي تستخدم لتسهيل التعامل مع Tensorflow. ثم استُعمِلت البيانات المولدة في الفصل الرابع دخلاً لهذه الشبكة العصبونية. في حالة تدريب الشبكة العصبونية فإنّ للشبكة ثلاثة مداخل الأول هو ناتج خرج تحويل CQT المُعالَج، والثاني والثالث هما ملفا بداية النغمة والنغمة الموسيقية ويمثلان الخرج الصحيح. أمّا في حالة الشبكة العصبونية المدرّبَة فإن الدخل يقتصر على خرج تحويل CQT وسيكون خرج الشبكة العصبونية في هذه الحالة هو مصفوفتان ثنائيتان تمثلان بداية النغمات الموسيقية وامتداد النغمات. ويمثل الشكل (5-1) بنية النظام ككل بشكل مبسط.

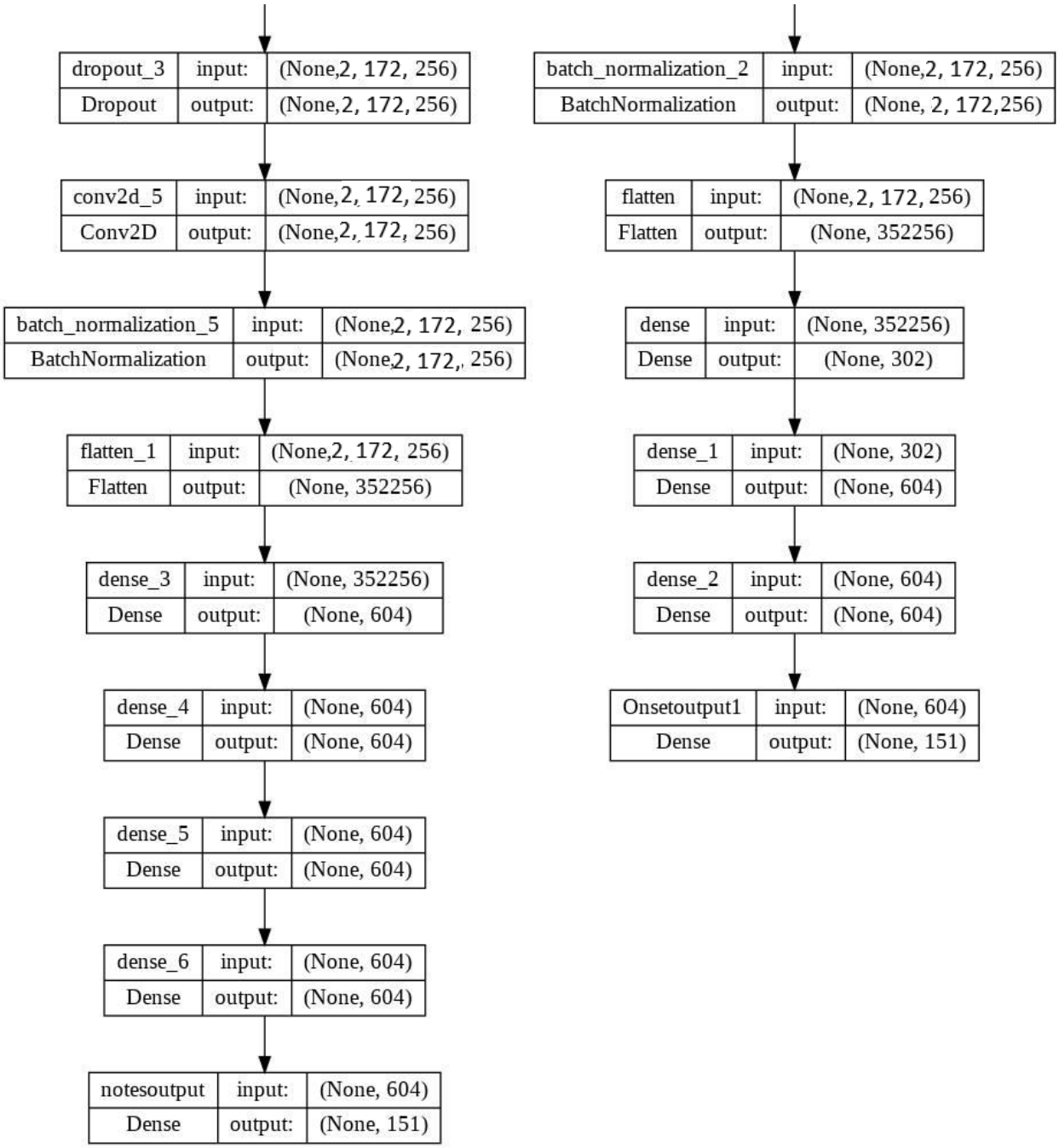


الشكل (5-1) البنية العامة للنموذج المصمم (الباحث)

5-1-1 بناء الشبكة العصبونية:

اعتُمِدَ النموذج الوظيفي Functional model في Keras الذي يسمح ببناء نماذج معقدة ومتداخلة بدلاً من النموذج التتابعي Sequential model (تكون فيه كل طبقة دخلا للطبقة التي تليها) بسبب تداخل الطبقات فيما بينها كون الدخل مشتركاً بين اكتشاف بداية النغمة الموسيقية ونهايتها. وعليه تم بناء النموذج كما هو موضح في االشكل (5-2) الذي يبين بنية الشبكة العصبونية واتصالات كلّ طبقة مع الطبقات المجاورة وحجم كل طبقة في الشبكة العصبونية وهو مولد باستخدام لغة python مكتبة Keras.

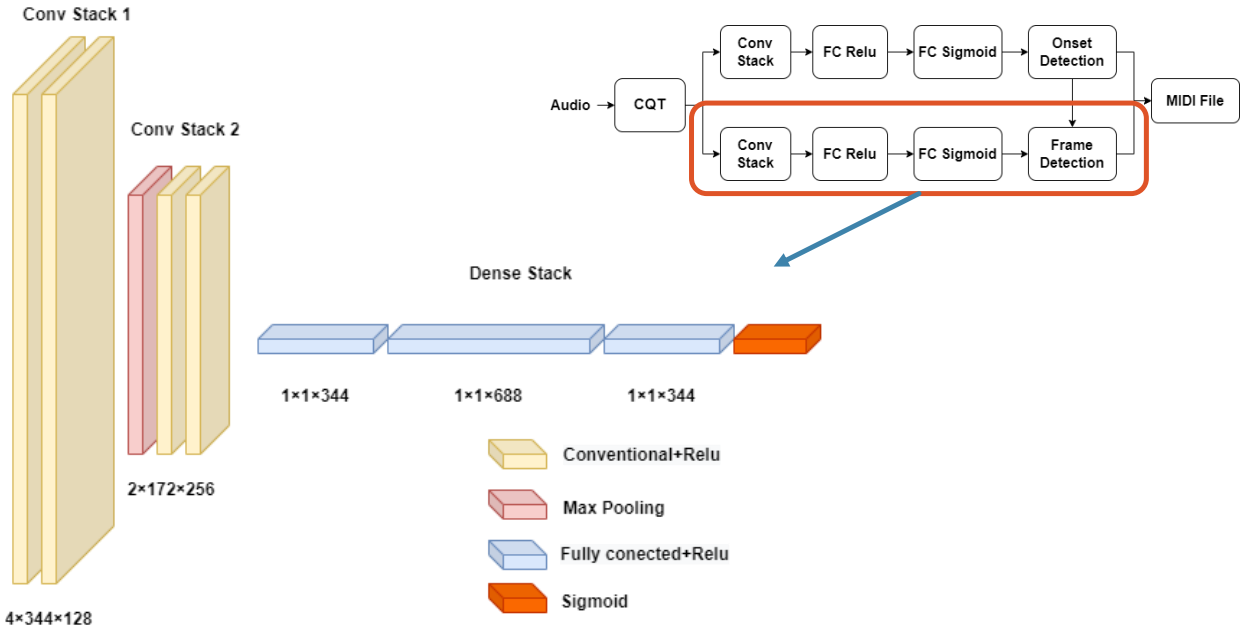

onsetinput1 | input: | [(None, 4, 344, 1)]
InputLayer | output: | [(None, 4, 344, 1)]
NoteConv1 | input: | (None, 4, 344, 1)
Conv2D | output: | (None, 4, 344, 128)
OnsetConv1 | input: | (None, 4, 344, 1)
Conv2D | output: | (None, 4, 344, 128)
dropout_1 | input: | (None, 4, 344, 128)
Dropout | output: | (None, 4, 344, 128)
batch_normalization | input: | (None, 4, 344, 128)
BatchNormalization | output: | (None, 4, 344, 128)
batch_normalization_3 | input: | (None, 4, 344, 128)
BatchNormalization | output: | (None, 4, 344, 128)
dropout | input: | (None, 4, 344, 128)
Dropout | output: | (None, 4, 344, 128)
conv2d_3 | input: | (None, 4, 344, 128)
Conv2D | output: | (None, 4, 344, 128)
conv2d | input: | (None, 4, 344, 128)
Conv2D | output: | (None, 4, 344, 128)
batch_normalization_4 | input: | (None, 4, 344, 128)
BatchNormalization | output: | (None, 4, 344, 128)
max_pooling2d | input: | (None, 4, 344, 128)
MaxPooling2D | output: | (None, 2, 172, 128)
dropout_2 | input: | (None, 4, 344, 128)
Dropout | output: | (None, 4, 344, 128)
conv2d_1 | input: | (None, 2, 172, 128)
Conv2D | output: | (None, 2, 172, 256)
max_pooling2d_1 | input: | (None, 4, 344, 128)
MaxPooling2D | output: | (None, 2, 172, 128)
batch_normalization_1 | input: | (None, 2, 172, 256)
BatchNormalization | output: | (None, 2, 172, 256)
conv2d_4 | input: | (None, 2, 172, 128)
Conv2D | output: | (None, 2, 172, 256)
conv2d_2 | input: | (None, 2, 172, 256)
Conv2D | output: | (None, 2, 172, 256)



الشكل (5-2) بنية الشبكة العصبونية من الدخل حتى الخرج مع أحجام الطبقات ونلاحظ أن الدخل مشترك ووجود خرجين (الباحث)

يبلغ عدد الوصلات الكلي بين الطبقات 12,776,916 وصلة.

يتخلل الطبقات المخفية عمليات انتخاب أعظمي Maxpolling وعمليات تخلص من الأوزان Droupout (تساعد في التغلب على مشكلة overfit) وعملية تقييس الرقعة Batch Normalization التي تساعد في تسريع عملية التعلم. كما يوضح الشكل (5-3) بنية كاشف امتداد النغمة الموسيقية الذي يحمل بنية مشابهة لكاشف بداية النغمة الموسيقية.



الشكل (5-3) بنية كاشف امتداد النغمة الموسيقية بدون طبقات Dropout و Batch normalization (الباحث)

5-1-2 توابع تنشيط الشبكة العصبونية:

إنّ عملية اختيار توابع التنشيط عملية تجريبية في كثير من الأحوال. ولكن لبعض أنواع المشاكل المعينة (Multi label classification في حالتنا) محددات عدّة ثابتة تعطي أداء أفضل كتابع الخسارة loss function وتابع تنشيط طبقة الخرج. يوضح الجدول (5-1) توابع تنشيط الشبكة العصبونية المستخدمة في دراستنا. يعد تابع ReLU الأفضل أداء في شبكات CNN عموماً كما أنّ تابع Sigmoid هو المستعمل في مشاكل التصنيف متعدد

المسميات Multi label classification وذلك بسبب مجال القيم من 0 حتى 1 لكلّ عصبونات الخرج التي تمثل كلاً منها علامة موسيقية واحدة (أي أنّ كلّ عصبون قيمته تقع فوق عتبة التنشيط يُعدُّ مفعلاً وهو ما يوافق تفعيل علامات موسيقية عدة في الوقت نفسه) بخلاف تابع Softmax الذي يحمل المجال نفسه من 0 حتى 1 ولكنه يَعُدُّ الخرج كتلةً واحدةً (مجموع عقد الخرج كلّها يساوي واحد والعقدة ذات القيمة الأكبر تعتبر مفعلة).

جدول (5−1) توابع التنشيط المستخدمة في الشبكة العصبونية (الباحث)

نوع الطبقة	تابع التنشيط
CNN	ReLU
Dropout	لا يوجد
Batch Normalization	لا يوجد
Dense (Full connected) طبقة مخفية	ReLU
Dense (Full connected) طبقة الخرج	Sigmoid

5−1−3 بارمترات الشبكة العصبونية:

تحسب قيمة تابع الخسارة المستخدم Loss Function بجمع تابعي الخسارة لكل من بداية النغمة الموسيقية وامتداد النغمة من نوع Binary cross−entropy ويعطى بالشكل التالي:

$$l_{total} = l_{onset} + l_{frame} \qquad (1-5)$$

ويعطى تابع الخسارة لبداية النغمة الموسيقية بالعلاقة التالية:

$$l_{onset} = \sum_{p=p_{min}}^{p_{max}} \sum_{t=0}^{T} CE(\mathrm{I}_{onset}(p,t), \mathrm{P}_{onset}(p,t)) \qquad (2-5)$$

إذ تمثل $p_{min/max}$ مجال العلامات الموسيقية من 0 حتى 154 في حين تمثل T عدد الإطارات في المقطوعة الموسيقية، ويمثل $\mathrm{I}_{onset}(p,t)$ تابع يعطي قيمة 1 عندما يكون هناك علامة موسيقية مفعلة في جدول الحقيقة في

العلامة رقم p والإطار رقم t ، بينما يمثل $\mathrm{P}_{onset}(p,t)$ القيمة التي توقعتها الشبكة العصبونية للعلامة رقم p والإطار رقم t، بينما تشير CE إلى تابع تنشيط Binary cross-entropy. يوضح الجدول (5-2) البارمترات المستخدمة في الشبكة العصبونية.

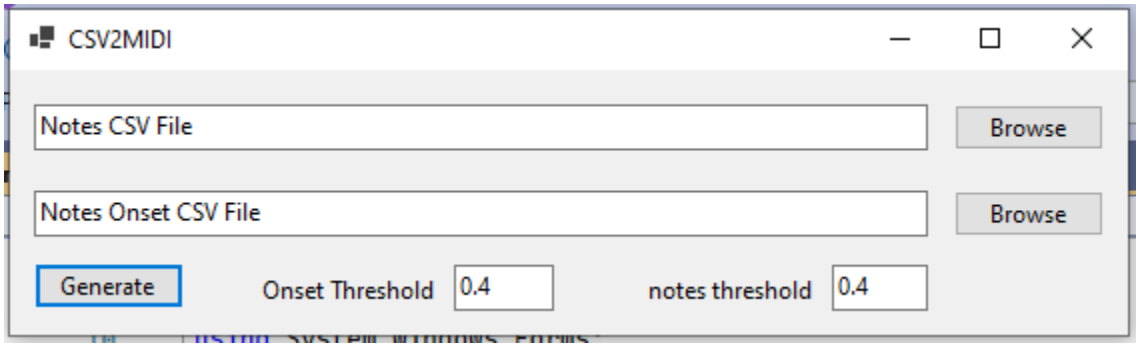
جدول (5-2) البارمترات المستخدمة في تصميم الشبكة العصبونية (الباحث)

المحدد	القيمة
تابع التنشيط	Binary cross-entropy
معدل التعلم	0.0001
المحسن (optimizer)	Adam optimizer
Batch size	16
عدد دورات التدريب Epochs number	10

5-1-4 تحويل خرج الشبكة العصبونية إلى ملف MIDI:

تعطي الشبكة العصبونية خرجين على شكل مصفوفة ثنائية الأبعاد، ولا يمكن ترجمة هذه النتائج إلى مقطوعة موسيقية دون معالجة هذين الخرجين. يمتلك خرج الشبكة العصبونية بنية مشابهة لملفات CSV المستعملة في التدريب التي شُرحت في الشكل (4-4). كما هو مذكور في [8] فإنّ بعض الإطارات تكون ذات أهمية أكثر من غيرها، ولا سيّما إطارات بداية النغمة الموسيقية وذلك لأنّ طبيعة الصوت تكون طرقية أو نقرية (ارتفاع مفاجئ وشبه آني في المطال) وتكون ذات مطال أعظمي وسرعان ما تسارع في الخفوت والدخول في طور امتداد النغمة. يجعل المطال الأعظمي من الأسهل اكتشاف بداية النغمة الموسيقية بدلاً من الاعتماد على بداية النغمة فقط لذلك قمنا بمنع كاشف امتداد النغمات من التنشيط في حال عدم تفعيل كاشف بداية النغمة الموسيقية قبله. لتحويل خرج الشبكة إلى ملف MIDI تمت كتابة برنامج بلغة C# مستخدمين مكتبة DryWetMIDI. بدايةً يتم انشاء قناتين الأولى للعلامات الغربية والثانية للعلامات الشرقية (يوضع فيها حدث انحناء نغمة بقيمة 6144) تُفحص العلامات الموسيقية كلّها في كلّ إطار من مصفوفة بداية النغمات بحثاً عن بداية العلامة الموسيقية. في حال اكتشاف بداية علامة موسيقية

يوضع حدث بداية علامة في ملف MIDI موافق لرقم العلامة المطلوبة في القناة الأولى في حال كانت العلامة غربية، وفي القناة الثانية في حال كانت شرقية. يوضع الحدث في الزمن الموافق لـ (رقم الإطار * 64*1000) مقاسا بالثانية. أمّا عند اكتشاف بداية علامة موسيقية يتم البحث في مصفوفة امتداد النغمات عن عدد الإطارات التي تمتد لها هذه النغمة ويوضع حدث نهاية العلامة الموسيقية برقم MIDI وقناة موافقة لبداية النغمة بقيمة زمنية مزاحة بمقدار (عدد الإطارات الممتدة * 64* 1000). يتم احتساب ما إذا كانت بداية النغمة أو امتداد العلامة الموسيقية مفعلتين عن طريق تحديد قيمة عتبة يدوية يمكن تغييرها (عادة ما تكون 0.5). يوضح الشكل (5-5) واجهة البرنامج المستخدمة في توليد ملف MIDI.



شكل (5-4) واجهة برنامج تحويل خرج الشبكة العصبونية إلى ملف MIDI (الباحث)

5-2 تقييم الأداء:

قمنا بتقييم النتائج وفق مقاييس معينة مشابهة لما هو متبع في الدراسات السابقة كي تكون الدراسة دقيقة قدر الإمكان وسنقوم بذكر وشرح آلية التقييم المتبعة.

5-2-1 معايير التقييم:

قمنا بتقييم أداء النظام لكل من العلامة الموسيقية والتردد الأساسي F0 وطول الإطار (بداية ونهاية النغمة معا) وذلك وفق المعايير التالية التي تستخدم عادة في تقييم مشاكل التصنيف من هذا النوع.

يوضح الجدول (5-3) الحالات الممكنة خرجاً للشبكة العصبونية وهو ما يعرف بمصفوفة التبعثر.

جدول (5-3) مصفوفة الارتباك وتمثل جدول الحقيقة لخرج الشبكة العصبونية

	التصنيف المتوقع		
التصنيف الحقيقي		Class = yes	Class = no
	Class = yes	True Positive	False Negative
	Class = no	False Positive	True Negative

بناء على جدول الحقيقة السابق يمكننا استخراج عدة معايير تعطى بالعلاقات الرياضية التالية:

1- الإحكام Precession: ويركز هذا المقياس على دقة النتائج الإيجابية ويقوم بحساب نسبة الحالات المتوقعة بنتيجة إيجابية صحيحة إلى عدد الحالات الكلي المتوقع نتيجةً إيجابيةً، التي قامت الشبكة العصبونية بتوقعها، ويعطى بالعلاقة:

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \qquad (3-5)$$

2- الحساسية أو الاسترجاع Recall: يركز هذا المقياس على نسبة الحالات الإيجابية المتوقعة الصحيحة على عدد الحالات الإيجابية الحقيقية الكلي ويعطى بالعلاقة:

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \qquad (4-5)$$

3- درجة F1 (F1 score): وهو متوسط موزون لكلّ من الإحكام والحساسية ويعطى بالعلاقة التالية:

$$F = \frac{2 \times P \times R}{P + R} \qquad (5-5)$$

إذ تشير N_{TP} إلى عدد الحالات الإيجابية الصحيحة، و N_{FP} إلى عدد الحالات الإيجابية الخاطئة، و N_{FN} إلى عدد الحالات السلبية الخاطئة، P تشير إلى الإحكام، وR إلى الحساسية، في حين F تشير إلى درجة F1.

5-2-2 آلية التقييم:

قمنا باستعمال مكتبة MIR_EVAL [33] لحساب كلّ من البارمترات السابقة. في حال تكرار علامتين موسيقيتين في كاشف بداية النغمات فإنّ النغمة الأخيرة تحسب توقعاً صحيحاً خاطئاً. تستخدم مكتبة MIR_EVAL معيارين لحساب هذه البارمترات:

- **المعيار الأول:** يتم اعتبار بداية العلامة الموسيقية صحيحا في حال كان الفرق الزمني بين العلامة المتوقعة والعلامة المرجعية ضمن مجال ±64ms فقط، على أن يكون فرق التردد بين العلامة المتوقعة والعلامة المرجعية لا يتجاوز ثمن المجال الترددي بين العلامة المرجعية الموافقة والعلامتين السابقة واللاحقة لهذه العلامة المرجعية (Note Only).
- **المعيار الثاني:** يتم اعتبار نهاية العلامة الموسيقية صحيحة في حال كانت نهاية العلامة ضمن المجال ±64ms عن نهاية العلامة المرجعية أو أن طول العلامة المتوقعة يمثل 80% من طول النغمة المرجعية الكلي، أيهما صحيح (Note With Offset) .

تحتاج مكتبة MIR_EVAL إلى صيغة معينة للدخل لتقييم النتائج لذلك وجب تحويل ملف MIDI الناتج إلى الصيغة الموضحة في الجدول (5-4). كُتب برنامج بلغة #C للقيام بعملية التحويل آنفة الذكر.

جدول (5-4) مثال لصيغة الدخل لمكتبة MIR_EVAL

رقم النغمة	زمن بداية النغمة (ثانية)	زمن النهاية (ثانية)	التردد (هرتز)
1	**0.447996**	**2.112033**	**311.128**
2	**0.895993**	**1.088**	**138.592**
3	**1.088**	**1.280007**	**231.245**
n	.	.	.

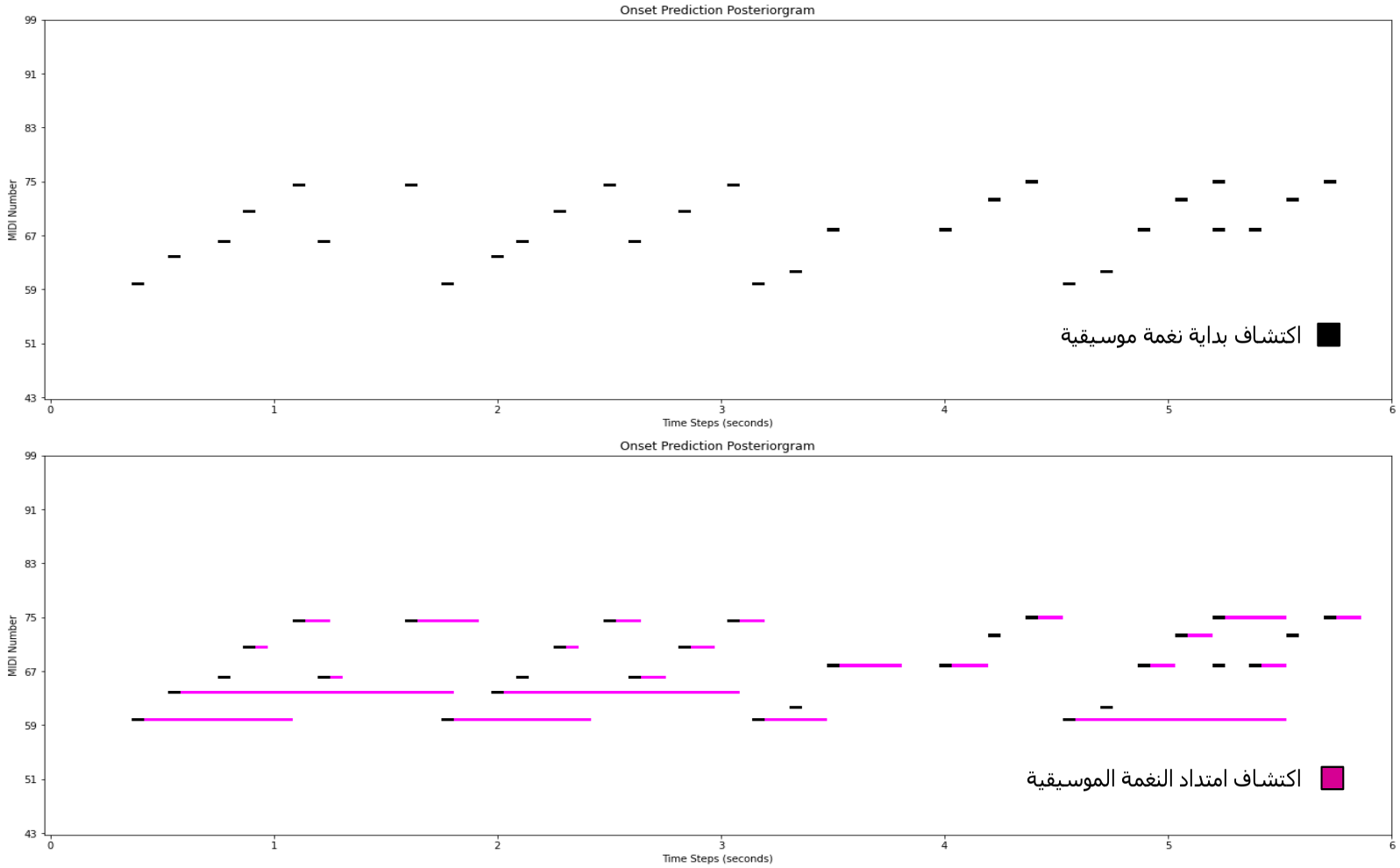
5-3 النتائج:

حُسبت نتائج التقييم بناءً على ما يزيد عن 30 دقيقة من المقطوعات الموسيقية الشرقية والغربية. استخدمنا جزءً من المقطوعات الموسيقية في مجلد MAPS Disklavier من قاعدة بيانات MAPS بوصفها عينات تقييم للموسيقا الغربية، في حين سُجّل عدد من المقطوعات الشرقية بواسطة بيانو الكتروني يدعم واجهة MIDI.

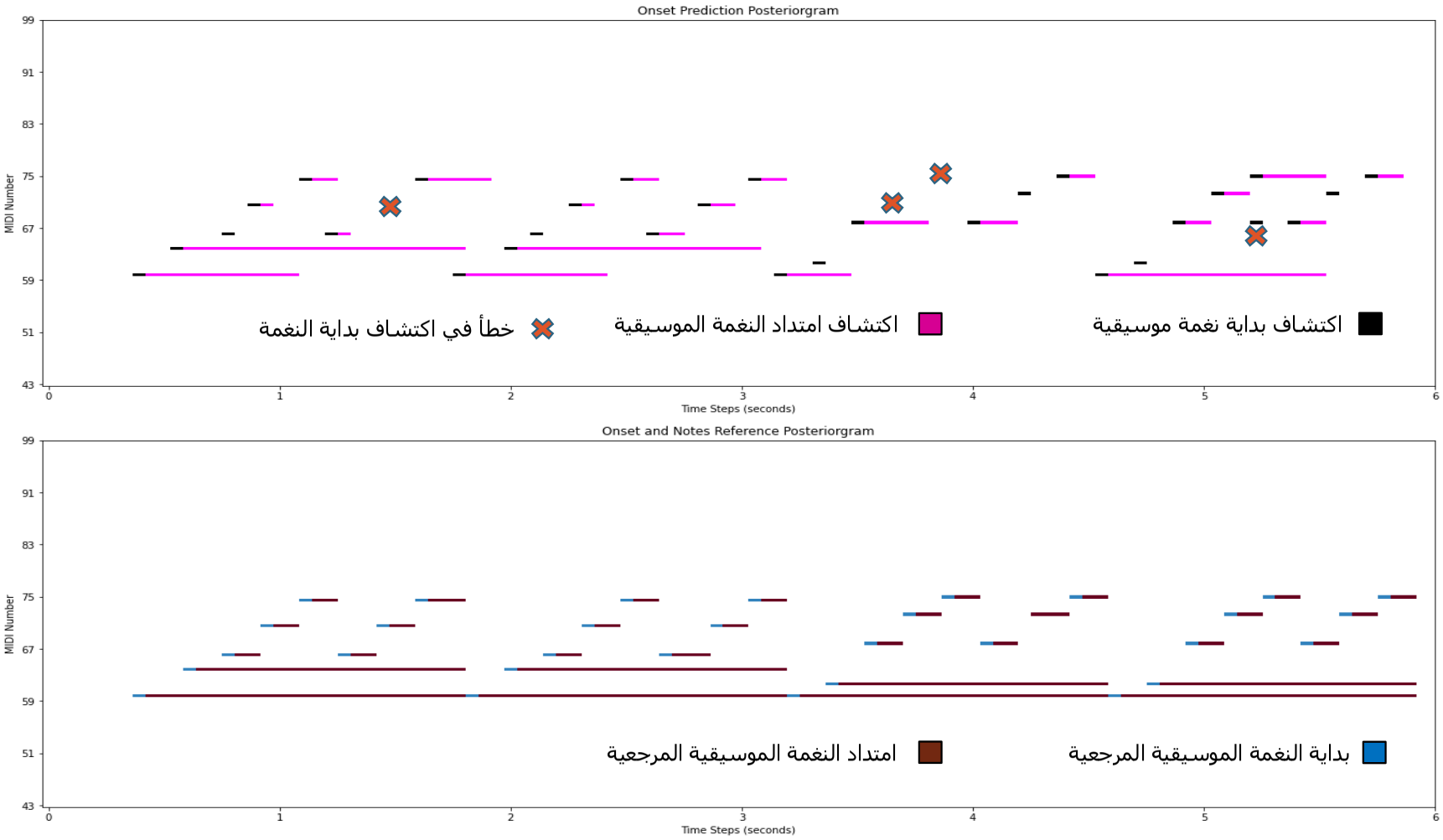
جدول (5-5) نتائج التقييم لكل من الإحكام (P)، والحساسية (R)، ودرجة F1 لكل من بداية العلامة الموسيقية فقط والعلامة الموسيقية مع نهاية إطار النغمة.

Note Only			Note With Offset		
P	**R**	**F1**	**P**	**R**	**F1**
94.40%	62.60%	75.28%	59.40%	39.39%	47.37%

كما يوضح الشكل (5-5) شكل لفات لبيانو لملف MIDI المستخرج وهو مولد باستخدام لغة برمجة Python.



شكل (5-5) لفات البيانو piano roll لملف MIDI المستخرج من الشبكة العصبونية لبداية النغمة فقط وهي موضحة باللون الأسود ، والإطار كاملا وهو موضح باللون الزهري (الباحث)



شكل (5–6) لفات البيانو piano roll لملف MIDI المستخرج من الشبكة العصبونية لكل من بداية النغمة وامتداد النغمة بالمقارنة مع ملف MIDI الأصلي مع إظهار الأخطاء في توقع بداية النغمة الموسيقية (الباحث)

5-4 مقارنة النتائج مع الدراسات السابقة:

تعدّ المقارنة الدقيقة مع الدراسات السابقة تحدياً حقيقياً وذلك بسبب تنوع الآلات الموسيقية، وتنوع آليات التقييم، ولكون معظم الدراسات السابقة لم تأخذ خطأ ربع العلامة بالحسبان. قُسّمت الدراسات إلى نوعين:

1- دراسات معنية بالموسيقا الشرقية والغربية على حد سواء وهي قليلة.

2 - دراسات معنية بالموسيقا الغربية فقط وهي الأكثر شيوعاً.

يمثل الجدول (5-6) والجدول (5-7) مقارنة بين الدراسات التي أخذت ربع العلامة بالحسبان والتي لم تأخذها على التتالي.

جدول (5-6) نتائج المقارنة لكل من الإحكام (P)، والحساسية (R)، ودرجة F1 لكل من بداية العلامة الموسيقية فقط والعلامة الموسيقية مع نهاية إطار النغمة للدراسات التي أخذت خطأ ربع العلامة بعين الاعتبار.

الدراسة	Note Only			Note With Offset		
	P	R	F1	P	R	F1
الدراسة الحالية	94.40%	62.60%	75.28%	59.40%	39.39%	47.37%
E. Benetos, 2015 [2]	53.22%	51.12%	51.91%	-	-	-

جدول (5-7) نتائج المقارنة لكل من الإحكام (P)، والحساسية (R)، ودرجة F1 لكل من بداية العلامة الموسيقية فقط والعلامة الموسيقية مع نهاية إطار النغمة للدراسات التي لم تأخذ خطأ ربع العلامة بعين الاعتبار.

الدراسة	Note Only			Note With Offset		
	P	R	F1	P	R	F1
الدراسة الحالية	94.40%	62.60%	75.28%	59.40%	39.39%	47.37%
[S. Sigtia 2016[34	44.97%	49.55%	46.581%	-	-	-
[R. Kelz 2016[35	44.27%	61.29%	50.94%	20.13%	27.80%	23.14%
C. Hawthorne 2018[2018]	84.24%	80.67%	82.29%	51.32%	49.31%	50.22%
A. Elowsson 2020 [11]	82.9%	85%	81%	64.6%	66.2%	63.1%

وفقاً لما توصلنا إليه بعد البحث فإنّنا نعتقد بإنّ النموذج المصمم يمثل حالة المعرفة في مجال استرجاع المعلومات الموسيقية والكتابة الآلية للموسيقا في الموسيقا الشرقية فقط بتحسين مقداره 28.54% في درجة F1 و 41.18% في الإحكام P، و 13.05% في الارتداد R. على الرغم من النتائج السابقة فإن النموذج أظهر نتائج أقل في حال إهمال خطأ ربع العلامة (أي في حالة الموسيقا غربية) مقارنة بالدراسات السابقة. نعزو ذلك إلى الأسباب التالية:

1- عدد العلامات الموسيقية في الموسيقا الشرقية ضعف عددها في الموسيقا الغربية تقريباً من ثمّ تضاعف عدد صفوف الخرج ليصبح (155).

2- بساطة النموذج المستخدم نسبيا لعدم توفر القدرة الحاسوبية الكافية لبناء نظام أكثر تعقيدا.

3- صعوبة الحصول على العينات الموسيقية الشرقية وعدم توفر مصادر سابقة لها.

5-5 الخاتمة Conclusion:

قمنا في هذا البحث بالتأكد من قدرة الشبكات العصبونية الالتفافية على استرجاع المعلومات الموسيقية والكتابة الآلية للموسيقا في الموسيقا الشرقية. بالاعتماد على التجارب التي قمنا بها والنتائج التي حصلنا عليها، وجدنا أنّ النموذج المصمم حسّن من الأداء على كل من صعيدي اكتشاف بداية النغمة الموسيقية وامتداد الإطار، وذلك عندما يكون خطأ ربع العلامة محسوباً. كما أنّنا وجدنا أنّ نقص عدد عينات التدريب، وبساطة النموذج النسبية، وزيادة عدد صفوف الخرج أثّرت سلباً على أداء النظام. كما أوضحنا كيفية إنشاء قاعدة بيانات تدريبية للموسيقا الشرقية بالاعتماد على قواعد البيانات الغربية والأدوات البرمجية المناسبة.

الآفاق المستقبلية

Future Work

نعتقد أنّ تضمين الشبكات العصبونية التكرارية طبقةً إضافيةً مع الشبكات العصبونية الالتفافية قد يسهم في تحسين النتائج. كما أنّ زيادة عينات التدريب لتشمل عدد أكبر من المقامات الموسيقية الشرقية وعدد أكبر من الآلات الموسيقية سيكون له أثر إيجابي على تحسين الأداء ككلّ ويزيد من الإمكانية العملية لتطبيقه. نقترح أيضاً توسيع النموذج لتحسين كل من اكتشاف بداية العلامة الموسيقية، وتردد العلامة، وامتداد العلامة. نوصي بتطبيق هذا النوع من التطبيقات في معاهد تعليم الموسيقا لما لها من أثر في تسهيل عملية التعليم واستخراج النوط الموسيقية لا سيّما للعازفين المبتدئين.

المراجع

References

[1] Klapuri, A., & Davy, M. (2006). Signal processing methods for music transcription. In Springer eBooks. https://doi.org/10.1007/0-387-32845-9.

[2] Benetos, E., & Holzapfel, A. (2015). Automatic transcription of Turkish microtonal music. The Journal of the Acoustical Society of America, 138(4), 2118–2130. https://doi.org/10.1121/1.4930187.

[3] Downie, J. S. (2008). The music information retrieval evaluation exchange (2005–2007): A window into music information retrieval research. *Acoustical Science and Technology*, *29*(4), 247–255. https://doi.org/10.1250/ast.29.247

[4] Maxwell, J. C. (1892) . *A Treatise on Electricity and Magnetism*. Oxford : Clarendon Press.

[5] Smaragdis, P., & Brown, J. C. (n.d.). Non-negative matrix factorization for polyphonic music transcription. *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No.03TH8684)*. https://doi.org/10.1109/aspaa.2003.1285860

[6] Raczynski, S., Vincent, E., Frédéric Bimbot, & Shigeki Sagayama. (2010). *Multiple Pitch Transcription using DBN-based Musicological Models.* 363–368.

[7] Nam, J., Jiquan Ngiam, Lee, H., & Slaney, M. (2011). A CLASSIFICATION-BASED POLYPHONIC PIANO TRANSCRIPTION APPROACH USING LEARNED FEATURE REPRESENTATIONS. *ISMIR*, 175–180.

[8] Hawthorne, C. (n.d.). Onsets and Frames: Dual-Objective Piano Transcription. *arXiv preprint arXiv*:1710.11153, 2017

[9] Wang, Q., Zhou, R., & Yan, Y. (2017). *A Two-Stage Approach to Note-Level Transcription of a Specific Piano*. *7*(9), 901–901.

[10] Goodman, T. T., & Batten, I. (2018). Real-Time Polyphonic Pitch Detection on Acoustic Musical Signals. *2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT),*. https://doi.org/10.1109/isspit.2018.8642626.

[11] Elowsson, A. (2020). Polyphonic pitch tracking with deep layered learning. *The Journal of the Acoustical Society of America*, *148*(1), 446–468.

[12] Alan, M. (2019), Middle Eastern Violin Method A Method for Teaching and Transcribing Middle Eastern Music. *Institutionen för Folkmusik*.

[13] Swift, A. (1997). A brief Introduction to MIDI. *Imperial College of Science Technology and Medicine*.

[14] Wolfe, J. (2019). *Note names, MIDI numbers and frequencies*. Unsw.edu.au. https://newt.phys.unsw.edu.au/jw/notes.html.

[15] *Quick Guide to Envelopes*. (n.d.). Making Music. https://making-music.com/quick-guides/envelopes.

[16] Lee, J.-H., Kang, J., Shim, W., Chung, H.-S., & Sung, T.-E. (2020). Pattern Detection Model Using a Deep Learning Algorithm for Power Data Analysis in Abnormal Conditions. *Electronics*, *9*(7), 1140. https://doi.org/10.3390/electronics9071140.

[17] Ye, Z. Muthu, K. Junichi, M. Takio, K. (2019). Music score estimation algorithm using octave hierarchy defined on logarithmic frequency. *International Workshop on Frontiers of Computer Vision (IW-FCV)At: South Korea*.

[18] Brown, J. C. (1991). Calculation of a constant Q spectral transform. *The Journal of the Acoustical Society of America*, *89*(1), 425–434. https://doi.org/10.1121/1.400476.

[19] de Cheveigné, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, *111*(4), 1917–1930. https://doi.org/10.1121/1.1458024.

[20] Johnson, N. S., Vulimiri, P. S., To, A. C., Zhang, X., Brice, C. A., Kappes, B. B., & Stebner, A. P. (2020). Invited Review: Machine Learning for Materials Developments in Metals Additive Manufacturing. *Additive Manufacturing*, 101641. https://doi.org/10.1016/j.addma.2020.101641.

[21] Dreyfus, S. (1962). The numerical solution of variational problems. *Journal of Mathematical Analysis and Applications*, *5*(1), 30–45. https://doi.org/10.1016/0022-247x(62)90004-5

[22] Gurucharan, M. (2020, December 7). *Basic CNN Architecture: Explaining 5 Layers of Convolutional Neural Network*. UpGrad Blog. https://www.upgrad.com/blog/basic-cnn-architecture/

[23] Meng, Z., & Chen, W. (2020). Automatic music transcription based on convolutional neural network, constant Q transform and MFCC. *Journal of Physics: Conference Series*, *1651*(1), 012192. https://doi.org/10.1088/1742-6596/1651/1/012192

[24] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," *2017 International Conference on Engineering and Technology (ICET)*, 2017.

[25] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386.

[26] Tahir Cagin, Chakrabarty, A., & Mannan, S. (2016). *Multiscale Modeling for Process Safety Applications*. https://doi.org/10.1016/c2011-0-07976-0.

[27] Dobroselsky, M. (2023, July 24). *melanchall/drywetmidi*. GitHub. https://github.com/melanchall/drywetmidi

[28] McFee, B., Raffel, C., Liang, D., Ellis, D., McVicar, M., Battenberg, E., & Nieto, O. (2015). Librosa: Audio and music signal analysis in Python. *Proceedings of the 14th Python in Science Conference*. https://doi.org/10.25080/majora-7b98e3ed-003

[29] SEEGER, C. (1958). PRESCRIPTIVE AND DESCRIPTIVE MUSIC-WRITING. *The Musical Quarterly*, *XLIV*(2), 184–195. https://doi.org/10.1093/mq/xliv.2.184.

[30] Yang, C., & Cong, G. (2019). *Accelerating Data Loading in Deep Neural Network Training*. https://doi.org/10.1109/hipc.2019.00037.

[31] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M. & others (2016). TensorFlow: A System for Large-Scale Machine Learning.. *OSDI* (p./pp. 265--283), .

[32] *keras-team/keras*. (2020, May 16). GitHub. https://github.com/keras-team/keras/

[33] Raffel, C., McFee, B., Humphrey, E. J., Salamon, J., Nieto, O., Liang, D. & Ellis, D. P. W. (2014). MIR_EVAL: A Transparent Implementation of Common MIR Metrics.. In H.-M. Wang, Y.-H. Yang & J. H. Lee (eds.), *ISMIR* (p./pp. 367-372), ..

[34] Sigtia, S., Benetos, E., & Dixon, S. (2016). An End-to-End Neural Network for Polyphonic Piano Music Transcription. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *24*(5), 927–939. https://doi.org/10.1109/taslp.2016.2533858.

[35] Kelz, R., Matthias Dorfer, Korzeniowski, F., Böck, S., Arzt, A., & Widmer, G. (2016). On the Potential of Simple Framewise Approaches to Piano Transcription. *ArXiv.org*, 475–481. https://doi.org/10.5281/zenodo.1416488.

Syrian Arab Republic

Ministry of Higher Education and Scientific Research

Damascus University

Faculty of Mechanical and Electrical Engineering

Department of Electronics and Communication Engineering

Musical Information Retrieval Using Digital Signal Processing and Deep Learning

A Thesis submitted in Partial Fulfillment of the Requirements for the Degree of Master in Applied Electronics Engineering

by

Eng. Ibrahim Wahbi

Supervisor

Dr. Hussein T. Bahbouh

Academic Year

2022-2023