



الجمهورية العربية السورية

وزارة التعليم العالي والبحث العلمي

جامعة دمشق

كلية الهندسة الميكانيكية والكهربائية

قسم هندسة الإلكترونيات والاتصالات

تطوير خوارزميات التعرف على الأشخاص باستخدام نماذج السير

أطروحة مقدمة لنيل درجة الدكتوراه في هندسة الاتصالات المتقدمة

إعداد

م. عبير صالح

المشرف المشارك

د. حازم ونوس

المشرف العلمي

د. طلال حمود

العام الدراسي

2025-2024

قرار لجنة الحكم وتوقيع اللجنة

أقدم أسمى آيات الشكر والامتنان والمحبة لأعضاء لجنة الحكم الموقرة لقبولهم قراءة خلاصة بحثي والحكم على رسالتي:

الأستاذ الدكتور محمد الحسين جامعة دمشق اختصاص اتصالات بصرية عضواً

الأستاذ المساعد الدكتور طلال حمود جامعة دمشق اختصاص معالجة إشارة عضواً

الأستاذ المساعد الدكتور جعفر سلمان جامعة طرطوس اختصاص نظم المعلومات عضواً

الدكتور رؤوف أبو حمدان جامعة دمشق اختصاص الرؤية الحاسوبية ومعالجة الصور الرقمية عضواً

الدكتورة جمان أبو جيب جامعة دمشق اختصاص اتصالات بصرية عضواً

بموجب قرار مجلس البحث العلمي والدراسات العليا رقم / 1328 / م.ب.د تاريخ 2025/1/13

والقاضي بتأليف لجنة الحكم.

قرار لجنة الحكم مع توقيع اللجنة

تصريح خطي Declaration

أنا الموقع أدناه، أصرح بعلمي الكامل وقولي:

1. أن كل ما ينتج عن البحث والأطروحة هو ملكية فكرية ومالية بالتشارك مع جامعة دمشق، وأنني ألتزم بأخذ موافقة الجامعة في حال رغبتني بنشر البحث أو جزء منها نصًا أو مضمونًا خارج إطار الجامعة (من دور نشر أو مكنتبات أو مواقع إلكترونية وغيرها من وسائل النشر).

2. أنه لا يوجد أي جزء من هذه الأطروحة مُقتَبَس من عمل آخر أو أنجز للحصول على شهادة أخرى في جامعة دمشق أو أية جامعة أو معهد تعليمي داخل أو خارج الجمهورية العربية السورية.

الاسم والتوقيع:

م.عبير صالح

شكر وتقدير

أقدم بخالص الشكر والتقدير للدكتور طلال حمود، المشرف الرئيس على هذا البحث ، على دعمه المستمر وتوجيهاته القيمة التي كانت نبراساً لي طوال مراحل البحث. لقد كان لخبرته الواسعة ورؤيته العميقة الدور الأكبر في إنجاز هذا البحث العلمي.

أتوجه بخالص الشكر والتقدير للدكتور حازم ونوس، المشرف المشارك على هذا البحث، على دعمه الكريم وإشرافه المتميز. لقد كانت توجيهاته العلمية وخبرته الواسعة عوناً كبيراً في تطوير هذا البحث.

الإهداء

إلى أمي

وإلى ابنتي سما

مع خالص حبي وامتناني

فهرس المحتويات List of Contents

i.....	قرار لجنة الحكم وتوقيع اللجنة
ii.....	Declaration تصريح خطي
iii.....	شكر وتقدير
iv.....	الإهداء
v.....	List of Contents فهرس المحتويات
xii.....	List of Tables فهرس الجداول
xiii.....	List Of The Terms and Abbreviations قائمة المصطلحات والاختصارات
xiv.....	المستخلص باللغة العربية
xv.....	English Abstract
1.....	الفصل الأول الإطار العام للبحث
1....	General Framework of the Research
2.....	1.1 مقدمة
4.....	2.1 بعض الدراسات المرجعية
4.....	1.2.1 تصنيف التعرف على المشي باستخدام أنواع مناهج المشي
9.....	2.2.1 تصنيف التعرف على المشي بالاعتماد على بيانات الإدخال
10.....	3.1 مشكلة البحث
12.....	4.1 هدف البحث
12.....	5.1 أهمية البحث
13.....	6.1 مسوغات البحث
13.....	7.1 محددات البحث
14.....	8.1 حدود البحث
14.....	9.1 منهج البحث وأدواته
15.....	10.1 بنية الأطروحة
17.....	الفصل الثاني مفاهيم أساسية في التعلم العميق والرؤية الحاسوبية
17.....	Basic Concepts in Deep Learning and computer vision
18.....	1.2 تعريف التعلم العميق

18	2.2 خلفية مصطلح عميق (deep).....
19	3.2 تاريخ التعلم العميق.....
20	4.2 مقارنة بين CPU و GPU و TPU.....
21	5.2 طريقة عمل الشبكات العصبونية الصناعية.....
22	6.2 أنواع الشبكات العصبونية الصناعية.....
23	7.2 مكونات شبكة التعلم العميق.....
24	8.2 أهم المكتبات المستخدمة في التعلم العميق.....
24	9.2 الخصائص النظرية المتعلقة بنموذج التعلم العميق.....
25	10.2 أهمية التعلم العميق.....
25	11.2 تحديات التعلم العميق.....
27	13.2 عيوب التعلم العميق.....
28	14.2 تطبيقات التعلم العميق.....
28	1.14.2 الرؤية الحاسوبية.....
29	2.14.2 معالجة اللغة الطبيعية.....
28	3.14.2 التعرف على الكلام.....
29	4.14.2 التعلم التعزيزي.....
29	15.2 التعلم العميق والرؤية الحاسوبية.....
29	1.15.2 تطور الرؤية الحاسوبية.....
31	2.15.2 طريقة عمل الرؤية الحاسوبية.....
32	3.15.2 أدوات برمجة الرؤية الحاسوبية.....
33	الفصل الثالث_الأسس النظرية لخوارزميات وتقنيات التعلم العميق المستخدمة.....
	Theoretical Foundations of Deep Learning Algorithms and Techniques
33	Used.....
34	1.3 مقدمة.....
34	2.3 الشبكة العصبونية التلافية.....
35	1.2.3 ميزات الشبكة العصبونية التلافية.....
35	2.2.3 تحديث الأوزان في الشبكة العصبونية التلافية.....
40	3.2.3 طبقات الشبكة العصبونية التلافية.....

44	4.2.3 استخراج الميزات.....
44	5.3 محول الرؤية (ViT) Vision Transformer (ViT).....
50	6.3 الشبكات العصبونية التكرارية (Recurrent Neural Network).....
52	1.6.3 الانتشار الخلفي عبر الزمن BPTT.....
54	2.6.3 المشكلات التي تواجه الشبكات العصبونية التكرارية.....
55	7.3 شبكة الذاكرة الطويلة قصيرة الأمد (Long Short Term Memory (LSTM).....
57	8.3 المعلومات والمعلومات الفائقة.....
58	9.3 استنظام الدفعة (Batch Normalization).....
59	10.3 الشبكة U-Net.....
60	11.3 تكثير الصورة.....
60	12.3 نقل التعلم Transfer Learning.....
62	13.3 البيانات المستخدمة في التعرف على المشي.....
62	1.13.3 البيانات المرئية (الصور والفيديو).....
63	2.13.3 البيانات الناتجة عن الحساسات.....
	الفصل الرابع تنفيذ خوارزميات التعلم العميق للتعرف على الأشخاص باستخدام بيانات المشي
68	المرئية.....
	Implementing Deep Learning Algorithms to Recognize People Using
69	Visual Gait Dataa.....
70	1.4 مقدمة.....
70	2.4 البيئة البرمجية.....
70	3.4 اختبار الشبكة العصبونية العميقة للحصول على أفضل بارامترات وتطبيق تقنية تكثير البيانات.....
70	1.3.4 طبقات الشبكة العصبونية المقترحة.....
72	2.3.4 تحسين النموذج.....
77	4.4 تنفيذ خوارزمية محول الرؤية باستخدام تقنية نقل التعلم.....
77	1.4.4 نموذج vit-base-patch-16-224.....
78	2.4.4 ضبط (fine tuning) نموذج ViT_base_patch_16_224.....
79	3.4.4 خوارزمية العمل المقترحة.....
80	5.4 مقارنة نتائج تطبيق CNN و ViT ومقارنة نموذج ViT مع بعض الأعمال المنشورة.....

81	6.4 ملخص نتائج الفصل الرابع
82	7.4 خاتمة الفصل الرابع
83	الفصل الخامس تنفيذ شبكات التعلم العميق باستخدام بيانات المشي الناتجة عن الحساسات
82	Implementing Deep Learning Networks Using Sensor Gait Data
84	1.5 مقدمة
84	2.5 البيئة البرمجية
85	3.5 تنفيذ شبكة U-Net للتمييز بين فترات المشي والراحة
85	1.3.5 وصف مجموعتي البيانات 7 و 8
86	2.3.5 الشبكة المقترحة لفصل عينات المشي والتدريب
90	3.3.5 نتائج التصنيف
90	4.5 تحسين خوارزمية CNN ودمجها مع تقنيات خوارزمية RNN
91	1.4.5 وصف مجموعات البيانات المستخدمة
92	2.4.5 تحليل مجموعة البيانات
93	3.4.5 اختبار الشبكة العصبونية التلافية CNN مع نوعين من توابع التفعيل
95	4.4.5 بنية نموذج CNN مع RNN المقترح
97	5.5 تحسين CNN ودمجها مع LSTM للتعرف على الأشخاص باستخدام عينات المشي مختلفة الطول
97	1.5.5 تحسين مجموعة البيانات OU-ISIR
99	2.5.5 هيكل النموذج المقترح للتعرف على الأشخاص باستخدام عينات مختلفة الطول
98	1.2.5.5 المكونات المخصصة squeeze and Excite
99	2.2.5.5 عنونة وتقسيم مجموعة البيانات
101	KerasTuner 3.2.5.5
104	6.5 ملخص نتائج تنفيذ خوارزميات التعلم العميق باستخدام بيانات الحساسات
	1.6.5 مناقشة تنفيذ شبكة U-Net باستخدام مجموعتي البيانات (7,8) التابعتان لمجموعة بيانات
104	whuGAIT
106	2.6.5 مناقشة اختبار CNN على حجوم دفعات مختلفة وتوابع تفعيل مختلفة
	3.6.5 مناقشة نتائج تنفيذ نظام IMU-GR باستخدام مجموعات البيانات (1,2,3,4) من مجموعة البيانات
108	whuGAIT
115	4.6.5 مناقشة نتائج تنفيذ نموذج التعرف على الأشخاص باستخدام عينات مختلفة الطول
116	7.5 مقارنة مع بعض الدراسات السابقة

119	الاستنتاجات العامة
120	الآفاق المستقبلية
121	الأبحاث المنشورة
121	References List قائمة المراجع

فهرس الأشكال List of Figures

6.....	الشكل (1-1) النقاط المهمة في هيكل الجسم المفصل وطول الأطراف
8.....	الشكل (2-1) مراحل استخراج الصور الظلية
10.....	الشكل (3-1) مراحل نظام التعرف التقليدي
10.....	الشكل (4-1) مراحل نظام التعرف الأتوماتيكي
36.....	الشكل (1-3) المخطط الصندوقي لنظام التعرف
39.....	الشكل (2-3) مجموع الدخل الموزون
40.....	الشكل (3-3) مداخل ومخرج عصبون في الشبكة العصبونية التلافية أثناء تحديث الأوزان
42.....	الشكل (4-3) تطبيق التلاف على الصورة
47.....	الشكل (6-3) البنية العامة للمحول المرئي
48.....	الشكل (7-3) مصنف المحول
50.....	الشكل (8-3) بنية كل من الشبكات العصبونية التكرارية (a) والشبكات العصبونية الأمامية (b)
51.....	الشكل (9-3) بنية الشبكة العصبونية التكرارية بعد فك الحلقات
54.....	الشكل (10-3) بنية LSTM
57.....	الشكل (5-3) أنواع التقييس
59.....	الشكل (11-3) بنية U-Net
69.....	الشكل (1-4) الطبقات الأساسية وتكرارها ضمن نموذج CNN المقترح
74.....	الشكل (1-4) بنية نموذج CNN المقترح لاختبار بيانات Market
74.....	الشكل (2-4) صورة من مجموعة البيانات Market التي سيطبق عليها تكثير الصور
75.....	الشكل (3-4) ناتج عملية تكثير الصور
76.....	الشكل (4-4) تقسيم الصور الى رقع
78.....	الشكل (5-4) دقة التدريب والاختبار الناتجة عن تطبيق نموذج ViT
84.....	الشكل (1-5) بنية U-Net المستخدمة
87.....	الشكل (2-5) ربط الطبقات المستخدمة في فصل عينات المشي عن الراحة
88.....	الشكل (3-5) تدريب النموذج على مجموعة البيانات
90.....	الشكل (4-5) الهيستوجرام لمجموعات البيانات 1,2,3,4
93.....	الشكل (4-5) البنية المقترحة لنموذج CNN مع RNN
96.....	الشكل (5-5) ترتيب عينات بيانات OU-ISIR قبل التعديل
97.....	الشكل (6-5) ترتيب عينات بيانات OU-ISIR بعد التعديل
99.....	الشكل (7-5) بنية النموذج المستخدم لاختبار عينات مختلفة الطول
100.....	الشكل (8-5) طبقة Squeeze and Excite
101.....	الشكل (9-5) الخوارزمية المتبعة في عنونة وتقسيم الداتا
103.....	الشكل (10-5) المخطط الصندوقي لعنونة وتقسيم البيانات
105.....	الشكل (11-5) خسارة الاختبار مع حجمي دفعة 512 و 256 وتابعي التفعيل Elu, ReLU
108.....	الشكل (12-5) منحنيات الدقة والخطأ عند تطبيق بنية النموذج CNN+ simpleRNN على مجموعتي البيانات 1,2

الشكل (5-13) منحنيات الدقة والخطأ عند تطبيق بنية النموذج CNN+ simpleRNN على مجموعتي البيانات 3,4	109
الشكل (5-14) منحنيات الدقة والخطأ عند تطبيق بنية النموذج CNN+ LSTM(64) على مجموعتي البيانات 2,1	110
الشكل (5-15) منحنيات الدقة والخطأ عند تطبيق بنية النموذج CNN+ LSTM (64) على مجموعتي البيانات	112
	4,3

فهرس الجداول List of Tables

69	 الجدول (1-4) تهيئة طبقة التلاف
70	 الجدول (2-4) تهيئة طبقات التجميع
70	 الجدول (3-4) تهيئة الطبقات المتصلة
70	 الجدول (4-4) تهيئة طبقة الإخراج
71	 الجدول (5-4) اختبار سرعة التعلم
72	 الجدول (6-4) اختيار أفضل قيمة لمعامل الإسقاط
73	 الجدول (8-4) قيم الدقة مقابل نواتي التجميع
78	 الجدول (9-4) مقارنة بين نموذجي CNN و ViT
79	 الجدول (10-4) مقارنة نتائج نموذج ViT مع بعض الأعمال المنشورة
68	 الجدول (1-5) الطبقات المستخدمة في فصل فترات الراحة عن المشي وبارامتراتهما
90	 الجدول (2-5) أكبر قيمة وأصغر قيمة لعدد العينات في مجموعات البيانات 1,2,3,4
91	 الجدول (3-5) المقارنة بين RELU و ELU
92	 نتائج تنفيذ CNN مع توابيع تفعيل وأحجام دفعات مختلفة
94	 الجدول (5-5) بنية طبقات CNN في نظام IMU-GR
94	 الجدول (6-5) طبقات RNN البسيطة و LSTM
94	 الجدول (7-5) نتائج تنفيذ نظام IMU-GR
102	 الجدول (8-5) قيم البارامترات المضبوطة باستخدام keras Tuner
115	 الجدول (9-8) المقارنة بين العمل المطروح والدراسات السابقة

List Of The Terms and المصطلحات والاختصارات

Abbreviations

Adaptive Momentum	ADAM	العزم التكيفي
Artificial Intelligence	AI	الذكاء الصناعي
Activation Function	AF	تابع التفعيل
Backpropagation Through Time	BPTT	انتشار الخلفي عبر الزمن
Batch Normalization	BN	طبقة استنظام الدفعة
Central Processing Unit	CPU	وحدة المعالجة المركزية
Convolutional Neural Network	CNN	شبكة عصبونية تلافئية
Convolutional Layer	CL	الطبقة التلافئية
Convolution	CONV	تلاف
Exponential Linear Unit	ELU	الوحدة الخطية الأسية
Fully Connected Layer	FC	الطبقة تامة الربط
Front-View Gait Dataset	FVG	مجموعة بيانات ذات الرؤية الأمامية
Gaussian Error Linear Unit	GELU	وحدة الخطأ الغاوسي الخطية
Gait Energy Image	GEI	صورة طاقة المشي
Global and Local Token	GLT	المتغيرات العامة والمحلية
Graphic Processing Unit	GPU	وحدة معالجة البيانات
Ground Reaction Forces	GRFs	قوى رد الفعل الأرضية
Gated Recurrent Unit	GRU	وحدة البوابات التكرارية
Image Augmentation	IA	تكثير الصورة
Inertial Measurement Unit	IMU	وحدة القياسات العطالية
Keep probability	Keep_prob	معامل الإسقاط
Learning Rate	LR	معدل التعلم
Long Short-Term Memory	LSTM	ذاكرة طويلة قصيرة الأمد
Multi-Layer Position-wise feedforward network	MLP	شبكة التغذية الأمامية متعددة الطبقات ذات الموقع الواحد
Multi headed Self-Attention	MSA	انتباه الذاتي متعدد الرؤوس
Pooling Layer	PL	طبقة الانتقاء
Rectified Linear Unit	ReLU	الوحدة الخطية المقومة
Recurrent Neural Network	RNN	شبكة عصبونية تكرارية
Tensor Processing Unit	TPU	وحدة معالجة الموترات
Vision Transformer	ViT	محول رؤية
Spatio-Temporal Attention Network	STGAN	شبكة الانتباه الزماني المكاني

المستخلص باللغة العربية

خلفية البحث: تم اتباع منهجية تجريبية للتعرف على الأشخاص بواسطة بصمة المشية، باستخدام الشبكة العصبونية التلافية (Convolutional Neural Network) و محول الرؤية (Vision Transformer) وبيانات مرئية من مجموعة البيانات Market، كما تم استخدام الشبكة U-Net لفصل فترات المشي عن الراحة بالاعتماد على بيانات خرج الحساسات المجدولة ويتضمن البحث أيضا دمج تقنية الشبكة العصبونية التلافية مع تقنيات الشبكة العصبونية التكرارية (Recurrent Neural Network) بنوعها الشبكة العصبونية التكرارية البسيطة (SimpleRNN) و الذاكرة الطويلة قصيرة الأمد (Long Short-Term memory) للتعرف على المشي باستخدام بيانات الحساسات المجدولة، وأخيرا استخدم نموذج تعلم عميق للتعرف على المشي باستخدام مجموعة البيانات OU-ISIR ذات العينات المختلفة الطول.

هدف البحث: تطوير خوارزميات التعرف على الأشخاص باستخدام نماذج السير وذلك من خلال استخدام مجموعة من خوارزميات التعلم العميق وتقنياته.

المواد والطرائق: استخدمت مكتبة التعلم العميق TFlearn لبناء نموذج الشبكة العصبونية التلافية واختيار أفضل بارامترات ومكتبة Pytorch لإعادة بناء نموذج محول الرؤية، واختبارهما والتحقق من دقتهما باستخدام بيئة Anaconda كما جرى بناء الشبكة U-Net ودمج نموذجي الشبكتين العصبونيتين التلافية و التكرارية ونموذج التعلم العميق المستخدم للتعرف باستخدام عينات مختلفة الطول باستخدام مكتبتَي Keras و Tensorflow واختبارها باستخدام بيئة google colab السحابية.

النتائج العملية: من أجل التعرف على الأشخاص باستخدام الصور تم اختبار قيم متعددة لبارامترات الشبكة العصبونية التلافية من أجل تحديد الدقة الأفضل، كما تم استخدام تقنية نقل التعلم للتعرف على الأشخاص عن طريق محول الرؤية ثم تم التحقق من الدقة للفصل بين فترات الراحة والمشي، وقد اختير أفضل نموذج ل الشبكة العصبونية التلافية بالنسبة لتتابع التفعيل واستنظام الدفعة ومن ثم دمج مع عدة أنواع من شبكات الشبكة العصبونية التكرارية وتبيان النموذج الأفضل. وبالنهاية تم تنفيذ نموذج تعلم عميق للتعرف على المشي باستخدام عينات مختلفة الطول، أظهرت النتائج دقة جيدة مما يجعل النماذج المقترحة نماذج متينة وقابلة للتطبيق في العديد من التطبيقات.

الاستنتاجات من وجهة نظر الباحثة: تقضي نتائج هذه الأطروحة إلى التحقق من فعالية النماذج المقترحة للتعرف على الأشخاص باستخدام مشيتهم.

الكلمات المفتاحية: الشبكة العصبونية التلافية، الشبكة العصبونية التكرارية، الذاكرة طويلة قصيرة الأمد، التعرف باستخدام المشي، محول الرؤية ، وحدة القياس العطالية.

المستخلص

English Abstract

The background of the research: This study explores the use of deep learning algorithms to recognize individuals based on their unique gait patterns. Specifically, Convolutional Neural Networks (CNN) and Vision Transformer (ViT) models were employed using visual data from the Market dataset. Additionally, the U-Net network was utilized to segment walking periods from rest periods using sensor output data from the whuGAIT dataset. The research also integrates CNN with Recurrent Neural Networks (RNN), including SimpleRNN and Long Short-Term Memory (LSTM), to identify individuals based on gait patterns derived from sensor data. Finally, a comprehensive model, combining several deep learning techniques, was developed to recognize gait using the OU-ISIR dataset, which includes samples of varying lengths.

The aim of the research: Enhancing person recognition algorithms through gait models by leveraging advanced deep learning algorithms and techniques.

Materials and Methods: The TFLearn deep learning library was used to construct and optimize the CNN model. PyTorch was employed to implement and test the Vision Transformer model, ensuring accuracy in an Anaconda environment. The U-Net network was designed for segmenting walking and rest periods, while the CNN+RNN models and the deep learning model for handling samples of different lengths were integrated using Keras and TensorFlow libraries. These models were tested within the Google Colab environment.

Practical Results: Multiple CNN configurations were evaluated to find the best-performing model for image-based recognition. Transfer learning was applied to improve the Vision Transformer's ability to recognize individuals, and the accuracy of distinguishing between walking and resting periods was validated. After selecting the optimal CNN model for activation functions and batch normalization, it was combined with various RNN models, leading to the identification of the most effective model. Finally, a deep learning model was developed to handle samples of varying lengths, achieving high accuracy. These results suggest the robustness and applicability of the proposed models in different practical scenarios.

Conclusions from researcher point of view: The results demonstrate the effectiveness of the proposed deep learning models for recognizing individuals based on their gait patterns. This work highlights the potential for these methods to be applied across a variety of applications.

Keywords: Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Gait Recognition, Vision Transformer (ViT), Inertial Measurement Unit (IMU).

الفصل الأول

الإطار العام للبحث

General Framework of the Research

الفصل الأول

الإطار العام للبحث General Framework of the Research

1.1 مقدمة

تعد الأنظمة التي تعتمد على التعرف الأوتوماتيكي للأفراد من أكثر التقنيات انتشارًا في التطبيقات الحديثة، خاصة في مجالات المراقبة الأمنية للمناطق الحساسة كالبنوك والمطارات. تعتمد هذه الأنظمة بشكل أساسي على القياسات الحيوية التي تُبرز الخصائص الفريدة لكل شخص استنادًا إلى سماته الفسيولوجية والسلوكية. وتشمل هذه التقنيات التعرف على القزحية، وبصمة الوجه، وبصمة الإصبع، وخط اليد، والمشية[1] .

تتميز الأنظمة الحيوية بدقتها العالية مقارنةً بالأنظمة التقليدية التي تعتمد على البطاقات أو كلمات المرور، إذ تستند إلى توقيعات حيوية مميزة لكل فرد، مما يجعلها غير قابلة للتكرار أو السرقة. وفي هذا البحث، نركز على استخدام توقيع المشية كأحد السمات السلوكية الفريدة للتعرف على الأفراد[2] .

يعرف توقيع المشية بأنه السمة الحيوية المستخلصة من أسلوب حركة الشخص، حيث تُعد طريقة المشي صفة فردية مميزة، مشابهة لبصمة الإصبع أو بصمة القزحية. تعتمد هذه السمة على عدة عوامل مثل وزن الشخص، وحركة الرأس، وطول الأطراف، وسرعة الحركة، والوضعية. وبالتالي، يمكن استخدام توقيع المشية كمعرف حيوي موثوق للتعرف على الأفراد بدقة وفعالية[3] .

بينما يعرف تحليل المشية على أنه قياس الكميات المسماة بمعلمات المشي الزمانية والمكانية (spatio_temporal parameter)؛ وترتبط هذه الكميات بخصائص الحركة البشرية[4] وهي المدة الزمنية للخطوة وطول مسافة الخطوة و قوة الخطوة و زوايا الحركة.

يعد تباين عوامل المشي هذا مؤشراً هاماً لتنوع التطبيقات المعتمدة على المشية كقياس بيومتری مثل أنظمة المراقبة والتحكم وكشف الجرائم ومجال الصحة [5].

في أعقاب جائحة كوفيد-19، برزت الحاجة إلى تعزيز تطبيق التباعد الاجتماعي بين الأفراد والمحيطين بهم كوسيلة أساسية للحد من انتشار العدوى. هذا الأمر أدى إلى تصاعد الطلب على خوارزميات تحديد الهوية البيومترية التي تتميز بالدقة، والأمان، والقدرة على العمل عن بُعد. وتُعد المشية واحدة من أبرز طرق القياسات الحيوية التي تلبي هذه المتطلبات، حيث تتميز بقدرتها على التعرف على الأشخاص من مسافات بعيدة وبدون تلامس، اعتماداً على أنماط ديناميكية مميزة يمكن تتبعها عبر كاميرات ثابتة متعددة [6].

مقارنة بوسائل القياسات الحيوية الأخرى مثل بصمة الإصبع، قزحية العين، التعرف على الوجه، أو تحليل الحمض النووي، تقدم تقنية المشية مزايا فريدة. فهي طريقة غير جراحية يصعب إخفاؤها أو تقليدها، ولا تتأثر بارتداء القفازات، الأقنعة، أو القبعات. بالإضافة إلى ذلك، تُظهر تقنية التعرف على المشية دقة وثباتاً حتى مع استخدام صور منخفضة الدقة، ولا تتطلب من الشخص أي تعاون أو انتباه أثناء عملية التعرف، مما يجعلها خياراً مثالياً للتطبيقات الحديثة.

نظراً لإمكانات ومزايا خوارزميات التعرف على المشي مقارنة بالقياسات الحيوية الأخرى، فإنها تُستخدم في مجموعة واسعة من تطبيقات الأمن المتعلقة بالمستشفيات ومراكز التسوق والبنوك والمنشآت العسكرية والمطارات والمؤسسات الدينية وتعتبر الميزة البيومترية الأكثر موثوقية في التحقيق الجنائي والوقاية من الجريمة والتعرف الجنائي لأنها أقل حساسية للجودة المنخفضة للصور مقارنة بالميزات البيومترية الأخرى [7].

تُعنى الدراسة المنهجية للحركة البشرية بقياس ووصف وتحليل الخصائص التي تميز حركات الإنسان. يُعتبر تحليل المشي أحد الأدوات الرئيسية لفهم مراحل المشي وتحديد المعلمات الحركية المرتبطة

بأحداث الحركة البشرية، بالإضافة إلى تقييم وظائف الجهاز العضلي الهيكلي بشكل كمي. نتيجة لذلك، وجد تحليل المشي تطبيقات واسعة في مجالات متعددة، مثل الرياضة، وإعادة التأهيل، وتشخيص الأمراض.

على سبيل المثال، في الرياضة، يُستخدم تحليل المشي لتحديد الأخطاء في أداء الرياضيين، مما يساعدهم على تحسين أدائهم. وفي مجال الصحة، يُستفاد منه لمراقبة تقدم شفاء المرضى، مثل استخدامه للتمييز بين الأفراد الأصحاء والمصابين بهشاشة العظام عبر تحليل بيانات مستشعرات التسارع والجيروسكوبات. كما تم تطوير تقنيات مراقبة الحركة لمتابعة حالات مرضى باركنسون.

في الهندسة الطبية الحيوية، يُعتبر تحليل المشي أداة أساسية لفهم الحركة البشرية بدقة، مما جعله موضوع اهتمام كبير للباحثين والأطباء على حد سواء، لدوره في تحسين فهمهم للوظائف الحركية وتطبيقها في العلاج والتشخيص.

2.1 بعض الدراسات المرجعية

إن العمل المقدم في هذه الأطروحة يركز على التعرف على الأشخاص بالاعتماد على مشيتهم باستخدام تقنيات وخوارزميات التعلم العميق المختلفة، كان هناك الكثير من الدراسات التي عالجت موضوع التعرف على الأشخاص باستخدام المشي، ويمكن تصنيف الدراسات السابقة إلى ثلاثة أنواع تبعاً للنهج المستخدمة وطريقة استخراج الميزات وأنواع البيانات المستخدمة.

1.2.1 تصنيف التعرف على المشي باستخدام أنواع مناهج المشي: يمكن تصنيف أنواع نهج

المشي إلى فئتين أساسيتين [8] :

• النهج المعتمدة على النموذج (model based approach):

ترتبط هذه النهج ارتباطاً وثيقاً بجسم الإنسان وحركاته، حيث تُستخدم لنمذجة الهيكل البشري وقياس السرعات والتسارعات باستخدام حساسات مثبتة على نقاط مختلفة من الجسم، مثل الهيكل

العظمي، القدمين، اليدين، والرأس، في حالة الاعتماد على الحساسات. أما عند استخدام الصور، يتم استخراج الهيكل المفصل للجسم (Skeleton) بشكل أكثر سهولة، حيث تعتمد الميزات عادةً على أجزاء مختلفة من الجسم. تشمل هذه الميزات أطوال الأجزاء (مثل الأطراف)، عرضها (مثل الرأس والجذع والركبتين والذراعين)، مواضع الرأس والكتفين، بالإضافة إلى المسارات الناتجة عن زوايا مفاصل مثل الورك.

النهج القائم على النمذجة أثبت فعاليته في التعرف على المشية بدقة من خلال تقنيات تقدير الوضعية (Pose Estimation) واستخراج سمات المشية. ومع ذلك، يواجه هذا النهج تحديات عدة، من أبرزها الانسدادات والظلال الموجودة في الصور، وصعوبة تحديد أجزاء الجسم بدقة من الصور الظلية الثنائية. لتحسين دقة التعرف، غالبًا ما يلزم تحويل الصور ثنائية الأبعاد عالية الجودة إلى نماذج ثلاثية الأبعاد، وهي عملية معقدة وتحتاج إلى موارد حوسبة كبيرة. بالإضافة إلى ذلك، فإن الصور الملتقطة بواسطة كاميرات المراقبة غالبًا ما تكون منخفضة الجودة، مما يؤدي إلى تقليل دقة التعرف على المشية ويُشكل تحديًا كبيرًا أمام التطبيقات العملية.

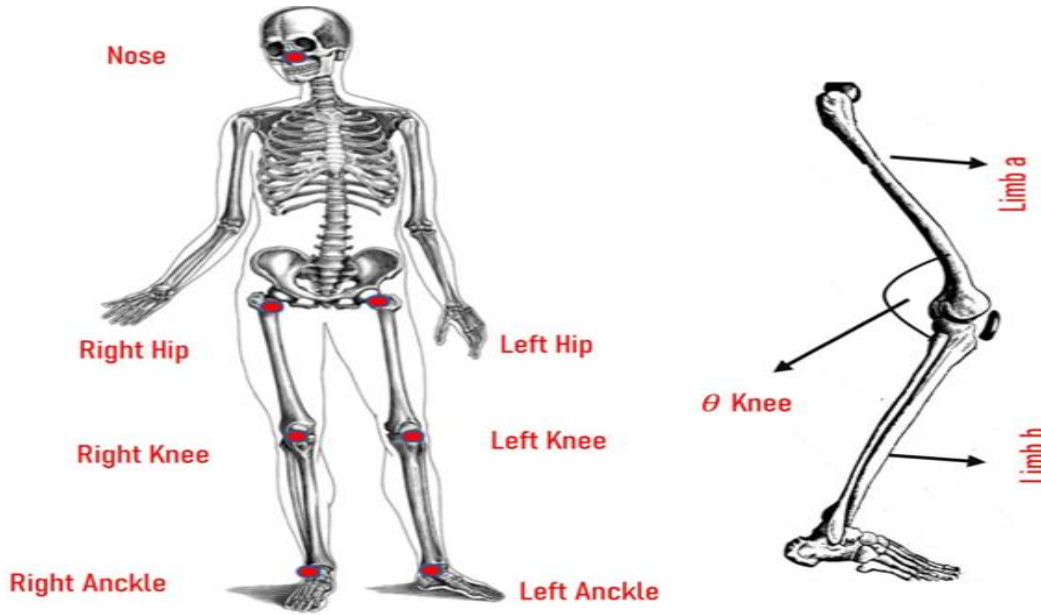
يعتمد استخراج ميزات المشي القائم على النموذج على تحديد مشية الإنسان من نمط دوران مفاصل الجزء السفلي من جسم الإنسان (الورك والركبة والكاحل). في الشكل (1-1) تم وضع الحساسات في عدة مفاصل، وتشير نقطة الأنف إلى النقطة المركزية لكلا الجانبين، بعد استخراج النقاط المهمة، يستخدم النموذج القائم أيضًا مسافات الأطراف الثابتة وزوايا المفاصل للتعرف على مشية الإنسان [9]. على سبيل المثال، قام الباحثون في [10] بنمذجة الهيكل العظمي البشري إلى 11 نقطة مفصالية ثم استخرجوا أربع مسافات ثابتة: طول الخطوة، ودرجة انحراف القدمين، ومسافة الكاحل الأيسر والأيمن، ومسافة الركبة اليسرى واليمنى. وعلاوة على ذلك، قاموا بحساب الزوايا الديناميكية، وهي زوايا الساق والورك، والساق والركبة، والساق والكاحل، وأخيرًا دمجوا السمات الثابتة والديناميكية،

كذلك اقترح الباحثون في [11] نهجاً قائماً على النموذج يسمى PoseGait، والذي قسم جسم الإنسان بالكامل إلى 18 نقطة مفصلية لتحديد المشية. كما قاموا باستخراج السمات الزمنية من الوضعيات

لتحسين معدل التعرف على المشية.

• النهج الخالية من النموذج (model free approach):

تركز الأساليب الخالية من النماذج على شكل الجسم أو حركة عملية المشي بأكملها ويتم فيها التعرف باستخدام أدوات مختلفة مثل استخراج صورة طاقة المشي (Gait energy image) GEI أو صورة تاريخ المشية من الصور الظلية المتاحة.



الشكل (1-1) النقاط المهمة في هيكل الجسم المفصل وطول الأطراف [8]

يعتمد إيجاد الصور الظلية على استخراج مشية الإنسان من الخلفية للتركيز على المنطقة المهمة للدراسة كبيانات إدخال خام، وذلك عبر تحليل مقاطع فيديو المشية إلى إطارات RGB، كل منها يحتوي على أجزاء من المقدمة والخلفية ومن ثم إنشاء الخلفية من الإطارات عن طريق حساب بعض العمليات

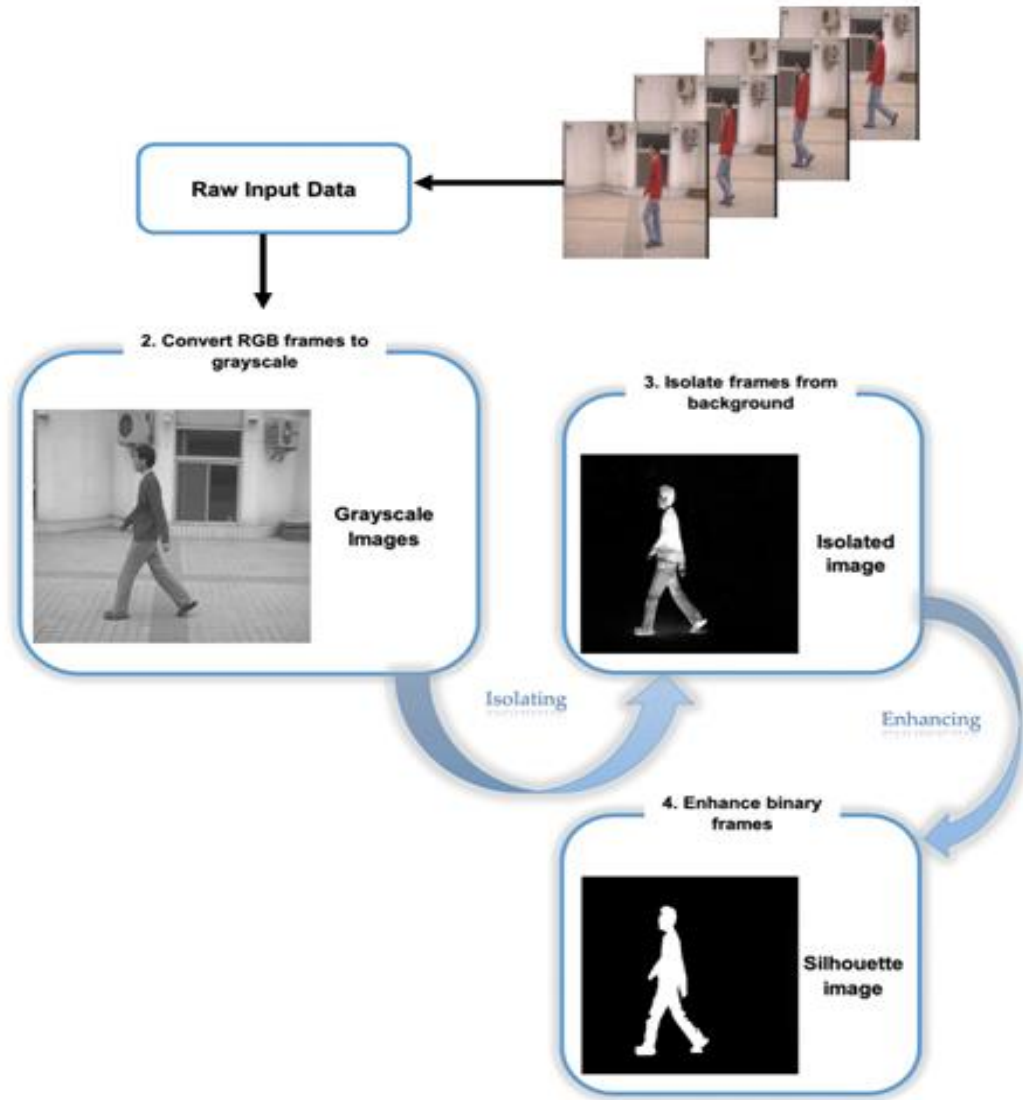
الحسابية، وهي التباينات والمتوسط. بعد ذلك تحدد مساحة الصورة الظلية عن طريق حساب المسافة بين البكسل والخلفية؛ بناءً على هذه المسافة، يتم تحديد ما إذا كان البكسل ينتمي إلى الخلفية أو المقدمة وأخيراً يتم إنشاء الصورة الظلية لتسلسل صور المشية حيث يوضح الشكل (1-2) مراحل استخراج الصورة الظلية من إطارات الفيديو. بشكل عام، يمكن للإجراءات الموضحة إنتاج صور ثنائية معزولة عالية الجودة، ومع ذلك، فإن بعض المشكلات تؤدي إلى أخطاء التجزئة [12] مثل الظلال وعتبة الخلفية/المقدمة؛ والكائن المتحرك في الخلفية. قبل استخدام هذه البيانات كمدخلات، يجب تطبيق العديد من إجراءات المعالجة المسبقة لتحسينها لتكون مناسبة لأغراض التعرف على الصور وتصنيفها على سبيل المثال اقترح الباحثون في [13] نموذجًا يستخرج مجالين، ميزات مكانية وزمنية، من صور الظلال تسمى شبكة الانتباه المكاني الزمني (STGAN). يمكن لهذا النموذج ملاحظة العلاقة بين الإطارات واكتشاف الاختلافات في المجال الزمني لتقليل الأخطاء التي تؤثر على معدل التعرف على المشية.

في [14] اقترح الباحثون خوارزمية للتغلب على عيوب تباين زاوية الرؤيا للمشاهد الملتقط من قبل كاميرات متعددة يعتمد السيناريو المقترح على تحويل الصور إلى صور ثنائية يتم التعرف على العناصر المتحركة والثابتة ومن ثم استخدام طرق الطرح لاستخراج الصور الظلية لكل شخص ومن ثم يتم استخراج صورة طاقة الحركة GEI لكل تسلسل، تقع زاوية الرؤية ضمن نطاق [0-270].

النهج الخالية من النموذج أقل حساسية لجودة الصورة وتكلفتها الحسابية أقل على الرغم من أنها ليست قوية بطبيعتها تجاه التباين والتدرج فهي تتأثر بتغير الملابس وحمل الأشياء وتغير زاوية الصورة وقد قدمت العديد من الدراسات للتغلب على هذه القيود فمثلاً في [15] كان الهدف فصل سمة المشي عن المظهر المرئي للشخص الذي يمشي (الصور تتأثر بألوان الملابس مثلاً) وحذف المعلومات الزائدة عن الحاجة من خلال اقتراح شبكة تعتمد على ترميز تلقائي، GaitNet، لفصل سمات الموضع عن المظهر المرئي بشكل صريح واستخدام LSTM متعدد الطبقات للحصول على سمة مشية مجمعة، ومن ثم جمع قاعدة بيانات مشية للمنظر الأمامي، تسمى FVG، بما في ذلك الاختلافات المختلفة لزوايا الرؤية،

وسرعات المشي، والحمل، وتغييرات الملابس، والخلفية والفجوات الزمنية.

يقسم النهج الخالي من النموذج أيضا إلى نوعين: النهج القائم على الحركة المتسلسلة (الصورة) والنهج القائم على الحركة المكانية الزمنية (spatio-temporal). في مناهج الحركة المتسلسلة، يتم تمثيل المشية كتسلسل للوضعيات البشرية، بينما يمثل النهج المكاني الزمني المشية من خلال رسم خريطة لتوزيع الحركة عبر المكان والزمان.



الشكل (1-2) مراحل استخراج الصور الظلية [8]

2.2.1 تصنيف أنواع التعرف على المشي بالاعتماد على بيانات الإدخال: تصنف طرق تحليل

المشية إلى ثلاثة أنواع بناءً على طرق الاستشعار المستخدمة لإجراء القياسات:

- **الحساسات اللاسلكية:** التي تستخدم مقاييس السرعة والتسارع للجسم وقوة الضغط الناتجة عن عملية ضغط القدم حيث تُستخدم سمات قوى رد الفعل الأرضية (GRFs) لتحليل المشية، هذه البيانات غالباً ما تكون مجدولة ومرتبطة زمنياً ويتم اتباع الخوارزميات التي تلتقط الميزات الزمنية في التعرف. هناك العديد من الدراسات التي استخدمت بيانات الحساسات للتعرف على المشي فمثلاً في [16] استخدم الباحثون هاتفاً ذكياً ومستشعراً بالقصور الذاتي يمكن ارتداؤه موضوعاً على كاحل الشخص للحصول على بيانات التسارع للتعرف على أنشطة المشي. تم تقسيم إشارات التسارع بناءً على الاكتشاف التلقائي للخطوات (دورات المشي). تم استخراج متجه الإشارات المجزأة، واستخدم هذا المتجه للتدريب

- **الصورة التي تعتمد على العمق (depth):** حيث يتم التقاط الصور عبر كاميرا تقيس العمق إلى الهيكل الداخلي للجسم مثل Kinect. وتم يتم استخدام معلومات الهيكل العظمي (skeleton) ثلاثية الأبعاد وتباين مسار نقاط المفاصل أثناء المشي أو ما يسمى التتبع الزمني لنقاط الهيكل العظمي لتحليل المشية. يتم حساب مقياس التباين بين مسارات نقاط الهيكل العظمي هذه وتشكل مصفوفات التباين نموذج المشية.

- **الصورة القياسية:** ويدخل ضمن هذا التصنيف كلا من الفيديو والصور، حيث يعتبر الفيديو سلسلة من الصور المرتبطة زمنياً، بعد أن يتم تقطيعه إلى العديد من الأطر، يجب مراعاة الترابط الزمني عندما تكون المدخلات بيانات فيديو ولا يلزم ذلك عندما تكون صور.

3.2.1 تصنيف أنواع التعرف على المشي بالاعتماد على طريقة استخراج الميزات: بشكل

استخراج الميزات أهم مرحلة من مراحل نظام التعرف باستخدام القياسات البيومترية جميعها وليس فقط المشية وتنقسم طريقة استخراج الميزات إلى فئتين:

- **طريقة استخراج الميزات التقليدية:** تعتمد هذه الطريقة على خوارزميات التعلم الآلي، حيث تتطلب العديد من التطبيقات عملية استخراج ميزات يدوية قبل مرحلة التدريب. يُعد هذا النهج ضرورياً في بعض

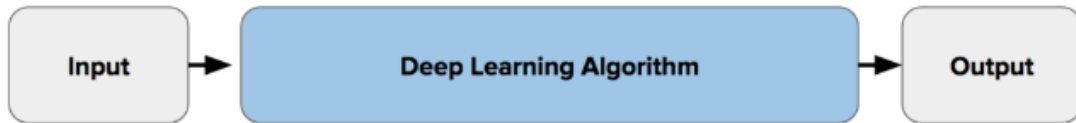
الحالات، مثل التطبيقات التي تعتمد على ذاكرة محدودة أو عند التعامل مع مجموعات بيانات صغيرة. في مثل هذه السيناريوهات، قد يؤدي استخدام طرق التعلم العميق إلى مشاكل مثل الإفراط في التدريب (Overfitting) بسبب قلة البيانات المتاحة وعدم تنوعها. في هذه الحالات، يُفضل استخدام خوارزميات تعتمد على استخراج الميزات يدويًا، حيث تُحدد الميزات الأكثر صلة وفعالية قبل التدريب، مما يقلل من تعقيد النموذج ويحسن أدائه في البيئات ذات الموارد المحدودة.

يبين الشكل (3-1) مراحل نظام التعرف في حال استخدام الميزات التقليدية.



الشكل (3-1) مراحل نظام التعرف التقليدي [18]

- طريقة استخراج الميزات الأوتوماتيكية: وفيها نستخدم استخراج الميزات باستخدام خوارزميات التعلم العميق، حيث أثبتت نماذج التعلم العميق، وخاصة الشبكات العصبونية التلافية (CNNs)، قدرتها على تعلم الميزات الهرمية تلقائيًا مباشرةً من البيانات الخام، متفوقة في كثير من الأحيان على الأساليب التقليدية التي تعتمد على الميزات المصنوعة يدويًا. يبين الشكل (4-1) مراحل نظام التعرف في حال استخدام الميزات الأوتوماتيكية.



الشكل (4-1) مراحل نظام التعرف الأتوماتيكي [18]

3.1 مشكلة البحث

تعاني الطرق التقليدية لتحديد هوية المستخدم مثل كلمات المرور ومعرفات المستخدم وبطاقات الشريط المغناطيسي ورموز PIN من العديد من العيوب، منها أنها تختبر صحة كلمة المرور أو رمز

PIN أو البطاقة المغناطيسية وليس المستخدم الفعلي، وبالتالي يمكن مشاركته بسهولة مع المستخدمين غير الشرعيين وهذا يشكل أمراً خطيراً على لأن نتيجة خرق الأمان لا تقتصر على أجهزة الكمبيوتر فقط، بل إنها قد تتيح الوصول إلى العديد من الموارد الأخرى مثل الحسابات المصرفية والسجلات الضريبية الشخصية والسجلات الطبية وبيانات الأمن القومي ونظام أمن المؤسسات والمطارات.

تعتبر القياسات الحيوية خياراً مثالياً لتحديد هوية المستخدم بسبب أن التحقق فيها يتم بواسطة الشخص نفسه، من أهم تطبيقات القياسات الحيوية هي أنظمة التحكم في الوصول فالتطبيقات مثل التحكم في الوصول إلى المباني، والتحكم في الوصول إلى الموارد الإلكترونية مثل أجهزة الكمبيوتر الشخصية وأجهزة الكمبيوتر المحمولة وأجهزة الصراف الآلي وأقوال الأمان الشخصية كلها تستفيد من تقنية القياسات الحيوية [9].

مقارنةً بطرق التعلم العميق، تواجه أساليب التعلم الآلي التقليدية العديد من التحديات والقيود التي تحد من كفاءتها في بعض التطبيقات. واحدة من أبرز هذه القيود هي محدودية عدد الميزات التي يمكنها معالجتها بفعالية، حيث تعتمد بشكل كبير على عملية استخراج الميزات يدوياً قبل التدريب. هذه الطريقة تجعلها أقل قدرة على التعامل مع البيانات المعقدة أو التعرف على الأنماط الخفية التي لا يمكن تحديدها بسهولة من قبل الإنسان.

إضافةً إلى ذلك، تعاني الطرق التقليدية من صعوبة في التعامل مع البيانات الكبيرة والمتنوعة، لأن تصميمها الأساسي لم يُهيأ لمعالجة الكميات الهائلة من البيانات أو التكيف مع اختلافاتها. عند التعامل مع بيانات ضخمة، تصبح النماذج التقليدية بطيئة وغير فعالة، وقد تتطلب معالجة مسبقة مكثفة.

في المقابل، تبرز تقنيات التعلم العميق كخيار مثالي للتغلب على هذه التحديات. فهي تتميز بقدرتها على استخراج الميزات تلقائياً عبر الطبقات العميقة في الشبكات العصبية، مما يسمح لها بتحليل البيانات الضخمة والتعامل مع تعقيداتها بشكل أكثر كفاءة ودقة. مع تطور الكاميرات التي تلتقط الصور

والفيديو والحساسات القادرة على التقاط الإشارات التي تعبر عن المشية ومع ظهور المعالجات القادرة على معالجة كميات ضخمة من البيانات وتطور تقنيات التعلم العميق يعتبر التعرف على المشي باستخدام تقنيات التعلم العميق أمراً مهماً لكثير من التطبيقات مثل أنظمة التحكم بالوصول وتحليل التشخيص الطبي والتحقق من الهوية وعلم الجريمة.

4.1 هدف البحث

إن الهدف الرئيسي لهذه الأطروحة هو تطوير خوارزميات التعرف على الأشخاص بالاعتماد على مشيتهم وباستخدام التعلم العميق وتطويرها لتحقيق أفضل دقة ممكنة، على وجه التحديد يتم التحقق من الخطوات البحثية التالية:

- التعرف على الأشخاص من خلال المشي بالاعتماد على بيانات صورية وخوارزمية الشبكة العصبونية التلافية CNN وتقنية تكثير الصور Image Augmentation.
- التعرف على الأشخاص من خلال المشي باستخدام خوارزمية محول الرؤية (Vision Transformer) وتقنية transfer learning.
- التمييز بين فترات الراحة والمشي بالاعتماد على معطيات مجدولة باستخدام شبكة U_Net.
- التعرف على الأشخاص باستخدام معطيات مجدولة بدمج خوارزميات CNN و RNN والمقارنة بينهم.
- التعرف على الأشخاص باستخدام معطيات مجدولة مختلفة الطول.

5.1 أهمية البحث

تبرز أهمية العمل في هذا المجال في جوانب عدة منها الأكاديمية البحثية ومنها التطبيقية والتجارية وخاصة بعد عام 2017 الذي شهد تطوراً في تفعيل التعلم العميق مما أدى إلى استخدام التعرف باستخدام القياسات البيومترية ومنها المشية بشكل واسع وكبير بعد أن أصبح الحصول على كميات ضخمة

من البيانات أمراً متاحاً.

إن التعرف على الأشخاص باستخدام المشي يعتبر مجالاً مهماً للدراسة لدى المختصين لما له من تطبيقات هامة يتفوق فيه على بقية الأنظمة البيومترية الأخرى مثل التحقيق الجنائي والوقاية من الجريمة. كما أن تحليل المشي ودراسه له تطبيقات مهمة جداً في المجال الصحي وفي المجال الأمني ولذلك تنوعت الأساليب المستخدمة في هذه الدراسات فالبعض يعتمد على استخدام التعلم الآلي والبعض الآخر يستخدم التعلم العميق وبعضهم يستخدم الصور الحية أو الفيديو ومنهم من يستخدم الصور الظلية أو الهيكل العظمي وحركته والبعض يستخدم معطيات الحساسات لتحقيق دقة أكبر في التعرف مع تكلفة منخفضة ولا يزال العمل مستمراً حتى الآن مع تنوع التطبيقات التي تستخدم المشية كقياسات بيومترية.

6.1 مسوغات البحث

لقد فرض الطلب المتزايد على الأنظمة البيومترية التي تستخدم التعلم العميق تحديات كبيرة في زيادة دقة هذه الأنظمة، وتطلب ذلك إجراء تجارب كثيرة للتوصل إلى الدقة الأمثل حيث تعتبر المشية من أهم القياسات البيومترية التي من السهل الحصول على مجموعة بيانات تمثلها وبالعودة إلى الدراسات المرجعية نجد أن التعلم العميق وخوارزمياته وتقنياته هي الأكثر استخداماً ولذلك سنستخدم هذه الخوارزميات في هذا البحث.

7.1 محددات البحث

إن أحد المحددات الأساسية في التعرف على الأشخاص هو استخراج الميزات حيث يوجد طريقتين في استخراج الميزات هي استخراج الميزات اليدوية واستخراج الميزات الديناميكية، لقد تم اعتماد استخراج الميزات الديناميكية في هذه الأطروحة حيث أن خوارزميات التعلم العميق تتيح ذلك باستخدام طبقاتها المتعددة، يتم استخراج الميزات بشكل هرمي للحصول في الطبقة الأخيرة على أفضل الميزات التي بواسطتها

نضمن الحصول على أفضل دقة. يتحدد هذا البحث من خلال:

- استخدام الشبكة العصبونية التلافية ومحول الرؤية لاستخراج الميزات والتعرف على الصور.
- استخدام شبكة u-net للفصل بين فترات المشي وفترات الراحة
- استخدام الدمج بين الشبكات العصبونية التلافية والشبكات العصبونية التكرارية للحصول على أفضل دقة في كل من المعطيات المجدولة المتساوية والمختلفة الطول.

8.1 حدود البحث

الحد المكاني: جرى تنفيذ هذا البحث في قسم هندسة الالكترونيات والاتصالات في كلية الهندسة الميكانيكية والكهربائية في جامعة دمشق.

الحد التقني: جرى استخدام مكتبات Tensorflow وTFlearn وPytorch وKeras المبرمجة باستخدام لغة python لبرمجة خوارزميات التعلم العميق.

-جرى استخدام بيئة Anaconda لتحقيق من دقة نماذج التعرف المعتمدة على البيانات الصورية المستخرجة من الفيديو.

-جرى استخدام بيئة google colab السحابية لتحقيق من دقة نماذج التعرف المعتمدة على البيانات المجدولة.

الحد الزمني: استغرق تنفيذ البحث مدة خمس سنوات.

الحد العلمي: جرى اتباع منهجية تجريبية في اختبار النماذج لتبيان النموذج ذو الدقة الأفضل.

9.1 منهج البحث وأدواته

تم اعتماد مراجع علمية من كتب ومقالات في مجالات علمية معتمدة في البحث العلمي كأدوات للبحث أثناء فترة التحضير للأطروحة، اعتمد في هذا البحث منهجية تجريبية في اختيار الخوارزميات

الأفضل والمناسبة للتطبيق على مجموعات المعطيات وفي إيجاد القيمة الأفضل للبارامترات والبارامترات الفائقة من أجل معايرة النماذج (fine tuning)، تم اختيار تقنيات مختلفة لإجراء التحسين على الدقة مثل تكثير البيانات ونقل التعلم، تم اختيار التعلم بإشراف لتدريب النماذج حيث استخدمت قوائم العنونة في حال وجودها مسبقا في مجموعة المعطيات أو في حال تمت عنونة مجموعات المعطيات غير المعنونة. تم اعتماد مكتبات التعلم العميق Tensorflow و TFlern و Pytorch و Keras والعديد من المكتبات الأخرى المبرمجة باستخدام لغة البرمجة python وتم اختيار بيئة anaconda و google colab لتنفيذ الأكواد وعرض النتائج. تعتبر البرمجيات المستخدمة من أهم البرمجيات التي تستخدم في هكذا أبحاث لما تؤمنه من بيانات عمل متكاملة تسهل للباحثين استخدام أهم مكتبات التعلم العميق بسهولة.

10.1 بنية الأطروحة

تتألف هذه الأطروحة من خمسة فصول:

تم في الفصل الأول تقديم الإطار العام للبحث والذي تضمن الحوافز وراء العمل في هذه الأطروحة والغرض من هذا العمل، كما عرض الفصل الأول بعض الدراسات المرجعية حول التعرف على الأشخاص باستخدام المشي.

يتناول الفصل الثاني مقدمة عن التعلم العميق والرؤية الحاسوبية وميزاتها وعيوبها.

يعرض الفصل الثالث الأساسيات النظرية لخوارزميات التعلم العميق وتقنياته والتي تم تطبيقها في هذا البحث.

يعرض الفصل الرابع اختيار أفضل بارامترات للشبكة العصبونية التلافية من أجل تحسين دقة التعرف على الأشخاص، من أجل كل بارامتر تم تجريب عدة قيم وتدريب النموذج ومن ثم تثبيت القيمة الأفضل ومتابعة اختبار بقية البارامترات، كما تم تنفيذ تقنية تكثير البيانات وتبيان أثرها على زيادة الدقة، كما تم تنفيذ خوارزمية محول الرؤية ViT الذي يقسم الدخل إلى مجموعة من الأجزاء ويبين الترابط من خلال

اختيار نموذج مدرب مسبقا باستخدام تقنية نقل التعلم، وقد نفذت الخوارزميتين ViT و CNN . على نفس مجموعة البيانات فتمت المقارنة بين النتائج ومناقشتها والمقارنة مع بعض الأعمال المنشورة.

يقدم الفصل الخامس طريقة فصل فترات المشي عن فترات الراحة، ومن ثم اختبار دمج خوارزمية CNN مع خوارزمية RNN البسيطة وخوارزمية LSTM وتبيان النتائج الأفضل للتعرف على الأشخاص باستخدام معطيات الحساسات المدمجة في الهاتف الذكي IMU، ثم نقترح نموذج للتعامل مع مجموعة معطيات مختلفة في أطوال العينات حيث تم اقتراح طريقة لتقسيم مجموعة المعطيات وعنونتها وتطبيق تقنية Squeeze and Excite لاستخراج أهم الميزات من الدخل ومن ثم دمج خوارزميتي CNN و LSTM وتبيان النتائج ومناقشتها والمقارنة مع بعض الأعمال المنشورة.

الفصل الثاني

مفاهيم أساسية في التعلم العميق والرؤية الحاسوبية

An Overview of Deep Learning and Computer Vision

الفصل الثاني

مفاهيم أساسية في التعلم العميق والرؤية الحاسوبية

1.2 تعريف التعلم العميق

تعريف: التعلم العميق هو فرع من فروع التعلم الآلي يعتمد على بنية الشبكة العصبونية الصناعية والتي تحاكي الشبكات العصبونية للدماغ البشري حيث تستخدم هذه الشبكة طبقات من العقد المترابطة تسمى الخلايا العصبونية التي تعمل معًا لمعالجة البيانات المدخلة وتميرها عبر هذه الطبقات بحيث تستنتج الميزات الأعلى في الطبقة الأعلى اعتمادًا على الميزات الأدنى في الطبقة السابقة وبناء نموذج استنادًا للعلاقات المعقدة بين البيانات ومن ثم اختبار هذا النموذج [17].

يمكن أن نطلق على مجموعات التعلم الآلي اسم تعلم عميق إذا كانت تستخدم التقانات التالية:

- تُستخدم وحدات المعالجة غير الخطية لاستخراج ومعالجة السمات في سلسلة من عدة طبقات. يعمل مخرج الطبقة السابقة كمدخل للطبقة التالية.
- إنشاء تمثيل هرمي من الميزات من خلال استنباط خصائص المستوى الأعلى من خصائص المستوى الأدنى للمدخلات.

إن قدرة التعلم العميق على التعرف على الأنماط والشذوذ في كميات هائلة من البيانات الخام تسمح له بتقديم نتائج تحليلية دقيقة. على سبيل المثال، لدى أمازون أكثر من 560 مليون سلعة للبيع وأكثر من 300 مليون مستخدم مسجل. بدون الذكاء الاصطناعي، حتى جيش كامل من المحاسبين سيجد صعوبة في تتبع معاملاتها.

2.2 خلفية مصطلح عميق (deep)

مصطلح "عميق" في سياق التعلم العميق (Deep Learning) جاء لوصف الهياكل الشبكية التي تحتوي على

عدة طبقات مخفية بين طبقة الإدخال وطبقة الإخراج. هذه الفكرة تعكس حرفياً العمق في التصميم الهيكلي للشبكة العصبونية.

3.2 تاريخ التعلم العميق

بدأت رحلة التعلم العميق في أوائل القرن العشرين، حيث بادر علماء الأعصاب بدراسة تفصيلية لعملية التفكير والمعالجة في الدماغ البشري. وفي عام 1943، قدّم العالمان "وارن ماكولوك" و"والتر بيتس" نموذجاً رياضياً للخلايا العصبونية، الذي أصبح معروفاً بـ "الخلايا العصبونية الصناعية". كان هذا النموذج الأول من نوعه الذي يحاكي كيفية عمل الخلايا العصبونية في الدماغ، وقد وفر أساساً جوهرياً لفهم الشبكات العصبونية الصناعية.

وفي الخمسينيات من القرن الماضي، أخذ العلماء خطوة أخرى نحو التعرف على القدرات العصبونية للآلة، عندما طور عالم النفس "فرانك روزنبلات" ما يعرف بـ "آلة بيرسيبترون".

كانت هذه الآلة هي أول نموذج لشبكة عصبية اصطناعية، وقد كانت خطوة كبيرة في تطوير الذكاء الاصطناعي. باستخدام هذه الآلة، بدأ العلماء في استكشاف قدرات الآلات على استيعاب البيانات والتعلم منها، وهو ما ساهم في بناء الأسس لما سيأتي بعد ذلك من تطورات في مجال الذكاء الاصطناعي والتعلم العميق.

فترة الستينيات والسبعينيات شهدت تقدماً ملحوظاً في مجال التعلم العميق، وذلك؛ عندما تم تطوير العديد من الخوارزميات الرئيسية التي أسهمت في تطور هذا المجال بشكل كبير، بما في ذلك "التعلم العكسي" و"انتشار الخلف".

في عام 1998، طور "يان لوكون" "الشبكات التلافية العميقة" [18]، وهي تقنية تُستخدم للتعرف على الصور بطريقة فعالة ودقيقة، مما سمح بتطبيقات جديدة في مجالات مثل التصوير الطبي وتحليل الصور الفضائية. وفي عام 2006، طور كلاً من "جيفري هينتون" و"سالي ورس" ما يُعرف بـ "الشبكات العصبونية العميقة"، وهي تقنية متطورة تُستخدم في مجموعة متنوعة من التطبيقات بما في ذلك الترجمة الآلية والتعرف على الكلام.

تمتاز هذه الشبكات بالقدرة على استيعاب وتمثيل البيانات بشكل هرمي وتحقيق نتائج دقيقة ومفيدة في مجموعة واسعة من المجالات.

في عام 2012، حقق فريق من الباحثين بقيادة "جيفري هينتون" فوزًا ملفتًا في مسابقة "Net Image" للتعرف على الصور، وذلك باستخدام شبكة عصبية عميقة تُعرف بـ "AlexNet"، وهو إنجاز اعتُبر تحولًا كبيرًا في فهم واستخدام الذكاء الاصطناعي في مجال تحليل الصور.

وفي عام 2016 طوّر فريق من مهندسي جوجل برنامج للذكاء الاصطناعي يُعرف بـ "ألفا جو"، والذي أدهش العالم بقدرته على هزيمة بطل العالم في لعبة "Go"، وهي لحظة تاريخية في تطور التكنولوجيا والذكاء الاصطناعي، حيث أظهرت قدرة الآلة على تفوق البشر في مجال تحليل الألعاب الاستراتيجية المعقدة.

وبهذا يكون التعلم العميق قد اكتشف في بداية عام 1940 نظريًا ولكن التطبيق الفعلي له بدأ في عام 2015 لحل المشكلات وذلك بسبب توفر كميات هائلة من البيانات التي تعتبر الوقود الأساسي للتعلم العميق وتطوير معالجات لديها القدرة على تحليل هذه البيانات مثل GPUs وTPUs.

4.2 مقارنة بين CPU و GPU و TPU

(CPU) هي اختصار لـ Central Processing Unit، وحدة المعالجة المركزية وتعتبر بمثابة عقل الكمبيوتر فهي الجهاز الأساسي الذي ينفذ التعليمات الخاصة ببرامج الكمبيوتر، وتتولى وحدة المعالجة المركزية معالجة جميع العمليات الحسابية والمنطقية والتحكمية الأساسية، كما تتولى وحدة المعالجة المركزية معالجة وظائف الإدخال والإخراج للبرنامج.

GPU هي اختصار لـ Graphical Processing Unit تعني وحدة المعالجة الرسومية، وهي مدمجة في كل وحدة معالجة مركزية في شكل ما. لكن بعض المهام والتطبيقات تتطلب تصورًا مكثفًا لا تستطيع وحدة معالجة الرسومات المدمجة التعامل معه مثل التصميم بمساعدة الكمبيوتر، والتعلم الآلي، وألعاب الفيديو، والبث المباشر، وتحرير الفيديو، وعلم البيانات، يمكن تنفيذ المهام البسيطة المتمثلة في عرض الرسومات الأساسية

باستخدام وحدة معالجة الرسوميات المدمجة في وحدة المعالجة المركزية. أما بالنسبة للمهام الأخرى المتطورة، يتم تصنيع وحدة معالجة الرسوميات الخاصة.

عند القيام بمهام رسومية مكثفة، فيمكن الحصول على خوادم وحدة معالجة الرسوميات وذلك لتقليل التكلفة المادية حيث أن خوادم وحدة معالجة الرسوميات هي خوادم مزودة بوحدة معالجة رسوميات يمكن استخدامها عن بُعد لتسخير قوة المعالجة الخام لإجراء حسابات معقدة.

Tensor processing unit (TPU) عبارة عن دائرة متكاملة خاصة ، لتسريع حسابات الذكاء الاصطناعي وخوارزمياته. طورتها Google [19] خصيصًا للتعليم العميق باستخدام مكتبة TensorFlow. بدأت Google في استخدام TPU في عام 2015؛ ثم قامت بنشره في عام 2018 يمكن الحصول على TPU كنسخة سحابية Cloud TPU. تتيح النسخة السحابية للمستخدم تشغيل مشاريع التعلم الآلي الخاصة على TPU باستخدام TF، النماذج التي كانت تستغرق أسابيع للتدريب على وحدة معالجة الرسوميات أو أي أجهزة أخرى يمكن أن تخرج في ساعات مع TPU. حيث يتم استخدام TPU فقط لمشاريع TensorFlow من قبل الباحثين والمطورين.

TPU هو معالج مخصص لإطار عمل TensorFlow ML. يتم استخدامهما بشكل أساسي لتحقيق أداء فعال على مجموعات البيانات الكبيرة للنمذجة وتدريب النماذج المعقدة بكفاءة. إنها أسرع وأقوى مقارنة بوحدة معالجة الرسوميات، مما يجعلها مناسبة لنشر التعلم العميق على نطاق واسع. في حين أن وحدة معالجة الرسوميات GPU هو معالج عام يستخدم للتعلم العميق والحوسبة العلمية والرسومات. إنها ميسورة التكلفة وأقل قوة. لذا، فهي خيار جيد للمشاريع الصغيرة.

5.2 طريقة عمل الشبكات العصبونية الصناعية

تعتمد الشبكات العصبونية الصناعية على مبادئ بنية وعمل الخلايا العصبونية البشرية، وتُعرف أيضًا باسم الشبكات العصبونية. تتلقى طبقة الإدخال في الشبكة العصبونية الصناعية، وهي الطبقة الأولى، مدخلات من مصادر خارجية وتمررها إلى الطبقة المخفية، وهي الطبقة الثانية. يحصل كل عصبون في الطبقة المخفية على معلومات من الخلايا العصبونية في الطبقة السابقة، ويحسب مجموع الأوزان ثم ينقله إلى الخلايا العصبونية في

الطبقة التالية. هذه العصبونات موزونة، مما يعني أن تأثيرات المدخلات من الطبقة السابقة يتم تحسينها إلى حد ما من خلال إعطاء كل مدخل وزنًا مميزًا. ثم يتم تعديل هذه الأوزان أثناء عملية التدريب لتحسين أداء النموذج. توجد العصبونات الصناعية، المعروفة أيضًا باسم الوحدات، في الشبكات العصبونية الصناعية. تتكون الشبكة العصبونية الصناعية بالكامل من هذه العصبونات الصناعية، والتي يتم ترتيبها في سلسلة من الطبقات. تعتمد تعقيدات الشبكات العصبونية على تعقيدات الأنماط الأساسية في مجموعة البيانات سواء كانت الطبقة تحتوي على عشرات الوحدات أو ملايين الوحدات. عادةً، تحتوي الشبكة العصبونية الصناعية على طبقة إدخال وطبقة إخراج بالإضافة إلى طبقات مخفية [19]. تتلقى طبقة الإدخال البيانات والتي تحتاج الشبكة العصبونية إلى تحليلها أو التعرف عليها.

في الشبكة العصبونية الصناعية تامة الربط (fully-connected)، توجد طبقة إدخال وطبقة مخفية واحدة أو أكثر متصلة واحدة تلو الأخرى. تتلقى كل خلية عصبية مدخلات من الخلايا العصبونية في الطبقة السابقة أو طبقة الإدخال. يصبح خرج خلية عصبية واحدة مدخلًا للخلايا العصبونية الأخرى في الطبقة التالية من الشبكة، وتستمر هذه العملية حتى تنتج الطبقة الأخيرة خرج الشبكة، ثم بعد المرور عبر طبقة مخفية واحدة أو أكثر، يتم تحويل هذه البيانات إلى بيانات قيمة لطبقة الإخراج، أخيرًا، توفر طبقة الإخراج خرجًا في شكل استجابة الشبكة العصبونية الصناعية للبيانات الواردة.

ترتبط الوحدات ببعضها البعض من طبقة إلى أخرى في غالبية الشبكات العصبونية. لكل من هذه الروابط أوزان تتحكم في مدى تأثير وحدة على أخرى. تتعلم الشبكة العصبونية المزيد والمزيد عن البيانات أثناء انتقالها من وحدة إلى أخرى، مما ينتج في النهاية خرجًا من طبقة الإخراج.

6.2 أنواع الشبكات العصبونية الصناعية

- الشبكات العصبونية الأمامية (FNNs) هي أبسط نوع من الشبكات العصبونية الصناعية، مع تدفق خطي للمعلومات عبر الشبكة. كمثال عليها هي الشبكات العصبونية التلافية (CNNs) وهي مخصصة لمهام التعرف على الصور والفيديو. تتمتع الشبكات العصبونية التلافية بالقدرة على تعلم الميزات تلقائيًا من الصور، مما

يجعلها مناسبة تمامًا لمهام مثل تصنيف الصور واكتشاف الكائنات وتجزئة الصور.

- الشبكات العصبونية التكرارية (RNNs) هي نوع من الشبكات العصبونية القادرة على معالجة البيانات المتسلسلة، مثل السلاسل الزمنية واللغة الطبيعية. تتمتع الشبكات العصبونية التكرارية بالقدرة على الحفاظ على حالة داخلية تلتقط المعلومات حول المدخلات السابقة [20]، مما يجعلها مناسبة تمامًا لمهام مثل التعرف على الكلام ومعالجة اللغة الطبيعية وترجمة اللغة.

7.2 مكونات شبكة التعلم العميق

شبكة التعلم العميق تتألف من عدة مكونات، وهذه المكونات تشكل هيكل الشبكة العصبونية العميقة:

- طبقة المدخلات (Input Layer) هذه الطبقة مسؤولة عن استقبال المدخلات من البيانات، كل وحدة في هذه الطبقة تمثل متغيراً أو ميزة من ميزات البيانات.
- الطبقات المخفية (Hidden Layers) تشكل هذه الطبقات الجزء الأكبر من الشبكة وتقوم بمعالجة المعلومات واستخراج الأنماط، كل طبقة تتكون من عدد من الوحدات النمطية والتي تسمى عصبونات (Neurons)، وكل وحدة متصلة بكل الوحدات في الطبقات التالية والسابقة.
- الوزن (Weight) يتمثل الوزن في الاتصالات بين الوحدات في الطبقات المخفية، يتم تعديل الوزن خلال عملية التدريب لتحسين أداء الشبكة.
- الانحياز (Bias) يضاف الانحياز إلى مخرجات الوحدات في الطبقات لتمكين الشبكة من تعلم وتمثيل الدالة العامة.
- توابع التفعيل (Activation Functions): تُضاف توابع التفعيل إلى مخرجات الوحدات لتحديد ما إذا كانت الوحدة نشطة أم لا، حيث تساهم هذه الوظائف في إضافة عنصر غير خطي إلى الشبكة، مما يتيح لها التعامل مع التفاعلات المعقدة.
- طبقة المخرجات (Output Layer) تُنتج هذه الطبقة نتائج النموذج. يعتمد تنظيم هذه الطبقة على نوع المشكلة، حيث يمكن أن تكون طبقة واحدة لمشكلات التصنيف وأكثر من طبقة لمشكلات الانحدار.

8.2 أهم المكتبات المستخدمة في التعلم العميق

- Theano
- TensorFlow
- Keras
- Pytorch
- Nltk

9.2 الخصائص النظرية المتعلقة بنموذج التعلم العميق

- **قوة الحوسبة: (Computational Power):** يتطلب التعلم العميق قدرًا كبيرًا من الحوسبة لتدريب النماذج على كميات ضخمة من البيانات. كلما كانت إمكانيات الأجهزة أقوى، زادت سرعة أداء النموذج. ولذلك، يُفضل استخدام معالجات الرسومات (GPU) أو معالجات متخصصة (TPU) لتسريع عملية التدريب.
- **السعة: (Capacity):** تعبر السعة عن عدد العصبونات والأوزان في النموذج، وهي تحدد قدرة النموذج على تعلم الأنماط المعقدة في البيانات. مع زيادة السعة، يمكن للنموذج تعلم معلومات أكثر ولكن مع خطر حدوث التعميم الضعيف إذا كانت السعة أكبر من اللازم.
- **التقارب (Convergence):** يشير إلى النقطة التي يتوقف عندها النموذج عن تحسين تنبؤاته، حيث يصبح معدل الخطأ ثابتًا ولا يتحسن بعد فترة من التدريب. التقارب يعكس مدى فعالية عملية التدريب وقدرة النموذج على الوصول إلى الحل الأمثل.
- **التعميم (Generalization):** يعني قدرة النموذج على نقل المعرفة التي تعلمها من بيانات التدريب إلى بيانات غير مرئية أو مجموعة اختبار. كلما كان النموذج قادرًا على التعميم بدقة أكبر، زادت فاعليته في العالم الحقيقي. بناء نموذج يتمتع بأعلى مستوى من التعميم هو الهدف الأساسي.

- **الإفراط في التدريب (Overfitting):** يحدث عندما يكون النموذج قادرًا على تقديم تنبؤات دقيقة على بيانات التدريب، ولكنه يفشل في التنبؤ بدقة على بيانات الاختبار. هذا يشير إلى أن النموذج قد تعلم تفاصيل غير هامة أو ضوضاء في بيانات التدريب، مما يعيق قدرته على التعميم.
- **الإقلال في التدريب (Underfitting):** يحدث عندما يكون النموذج بسيطًا جدًا بحيث لا يستطيع التقاط الأنماط المعقدة في البيانات، مما يؤدي إلى أداء ضعيف على كل من بيانات التدريب والاختبار. يحدث هذا غالبًا نتيجة تبسيط مفرط في الخوارزمية أو نقص في تمثيل البيانات.

10.2 أهمية التعلم العميق

- **تحسين أداء المهام المعقدة :** يساعد التعلم العميق في تنفيذ المهام المعقدة مثل التعرف على الصور، والترجمة الآلية، وتحليل البيانات الضخمة بشكل أكثر دقة وكفاءة.
- **الاستفادة من البيانات الضخمة:** يعتمد التعلم العميق على معالجة وتحليل البيانات بشكل فعال، مما يسمح باستخدام مجموعات بيانات كبيرة لتدريب النماذج وتحسين أدائها.
- **التعرف على الأنماط المعقدة:** من خلال استخدام طبقات أعمق في الشبكات العصبية، يتمكن التعلم العميق من استخراج وتمثيل الأنماط الدقيقة والمعقدة في البيانات.
- **تطبيقات متعددة المجالات:** يُستخدم التعلم العميق في مجموعة واسعة من التطبيقات مثل الطب (تشخيص الأمراض)، والتمويل (التنبؤ بالأسواق)، والتصوير الفضائي (تحليل الصور الفضائية)، والتكنولوجيا (المساعدات الشخصية).

11.2 تحديات التعلم العميق

- هناك تحديات وتعقيدات تواجه مجال التعلم العميق، ومن بين هذه التحديات:
- **الاعتماد على كميات ضخمة من البيانات:** يتطلب تحقيق أداء جيد في التعلم العميق وجود مجموعات بيانات

كبيرة، مما قد يشكل تحديًا في المجالات التي لا تتوفر فيها بيانات كافية.

- **تعقيد تصميم الشبكات العصبونية:** يتطلب تصميم هيكل الشبكة واختيار المعلمات الفائقة (Hyperparameters) معرفة عميقة وخبرة بالمجال. بعض النماذج تحتاج إلى هندسة معمارية متقدمة لضمان الأداء الأمثل.
- **استهلاك الموارد والوقت:** تدريب النماذج العميقة غالبًا ما يكون مكلفًا من حيث الزمن والموارد الحاسوبية، مما يجعل تحسين الأداء وإيجاد النموذج المثالي تحديًا كبيرًا.
- **صعوبة التفسير:** تعتبر النماذج العميقة غالبًا بمثابة "صندوق أسود"، حيث يصعب تفسير القرارات التي تتخذها، خاصة في المجالات التي تتطلب تفسيرات دقيقة وشفافة.
- **حساسية تجاه التغيرات في البيانات:** يمكن أن تؤثر التغيرات الطفيفة في البيانات أو التوزيع الإحصائي لها بشكل كبير على أداء النموذج، مما يستدعي عناية مستمرة بجودة البيانات.
- **مخاطر الأمن والهجمات:** يمكن استغلال نقاط الضعف في النماذج العميقة من خلال الهجمات الإلكترونية أو التلاعب، خاصة في البيئات المفتوحة عبر الإنترنت، مما يتطلب استراتيجيات حماية فعالة.

12.2 مزايا التعلم العميق

- **هندسة الميزات الآلية:** تتيح خوارزميات التعلم العميق اكتشاف الميزات ذات الصلة وتعلمها تلقائيًا من البيانات، دون الحاجة إلى عملية هندسة ميزات يدوية.
- **الفعالية مع كميات ضخمة من البيانات:** يتميز التعلم العميق بقدرته على التعامل مع البيانات الضخمة، مما يحسن الأداء بشكل كبير ويُمكن من تحليل مجموعات بيانات معقدة.
- **القدرة على فهم الأنماط المعقدة:** يساعد التعلم العميق في تمثيل وفهم الأنماط الدقيقة والمعقدة في البيانات، ما يجعله يتفوق في مهام مثل التعرف على الصور والفيديوهات.

- **التعلم الهرمي والتلقائي:** يمكن لنماذج التعلم العميق تعلم التمثيلات الهرمية بشكل تلقائي، مما يعزز استيعاب المفاهيم والأنماط العميقة
- **فهم التفاعلات المعقدة:** تساعد الهياكل العميقة في تمثيل العلاقات بين المتغيرات بشكل أفضل، مما يؤدي إلى تحسين أداء النماذج في المهام المختلفة.
- **قابلية التوسع:** يمكن لنماذج التعلم العميق التوسع بإضافة طبقات لمعالجة مجموعات بيانات أكبر وأكثر تعقيدًا، والتعلم من كميات هائلة من المعلومات.
- **المرونة:** يمكن تطبيق التعلم العميق على مجموعة واسعة من المهام والبيانات، بما في ذلك الصور والنصوص والكلام، مما يجعله أداة متعددة الاستخدامات
- **التحسين المستمر:** تتمتع نماذج التعلم العميق بإمكانية تحسين الأداء مع زيادة حجم وجودة البيانات المتاحة، مما يجعلها تتطور باستمرار مع الوقت.

13.2 عيوب التعلم العميق

- **متطلبات حسابية عالية:** تتطلب نماذج التعلم العميق كميات كبيرة من البيانات والموارد الحسابية للتدريب والتحسين.
- **كميات كبيرة من البيانات المصنفة:** غالبًا ما تتطلب نماذج التعلم العميق كمية كبيرة من البيانات المصنفة للتدريب، والتي قد تكون مكلفة وتستغرق وقتًا طويلاً للحصول عليها.
- **القدرة على التفسير:** قد يكون تفسير نماذج التعلم العميق أمرًا صعبًا، مما يجعل من الصعب فهم كيفية اتخاذها للقرارات.
- **الإفراط في التدريب:** يمكن لنماذج التعلم العميق في بعض الأحيان أن تتبع مسار بيانات التدريب، مما يؤدي إلى ضعف الأداء على البيانات الجديدة وغير المرئية.
- **طبيعة الصندوق الأسود:** غالبًا ما يتم التعامل مع نماذج التعلم العميق على أنها صناديق سوداء، مما يجعل من الصعب فهم كيفية عملها وكيف توصلت إلى تنبؤاتها.

14.2 تطبيقات التعلم العميق

1.14.2 الرؤية الحاسوبية: من خلال الرؤية الحاسوبية، يمكن لنماذج الذكاء الاصطناعي للتعلم العميق تمكين الآلات من تحديد وفهم البيانات المرئية [21]. تتضمن بعض التطبيقات الرئيسية للتعلم العميق في الرؤية الحاسوبية ما يلي:

اكتشاف الكائنات والتعرف عليها: يمكن استخدام نموذج التعلم العميق لتحديد وتحديد موقع الكائنات داخل الصور ومقاطع الفيديو.

تصنيف الصور: يمكن استخدام نماذج التعلم العميق لتصنيف الصور إلى فئات مثل الحيوانات والنباتات والمباني وتصنيف الصور الطبية.

تجزئة الصور: يمكن استخدام نماذج التعلم العميق لتجزئة الصور إلى مناطق مختلفة، مما يجعل من الممكن تحديد ميزات معينة داخل الصور.

2.14.2 معالجة اللغة الطبيعية (NLP): يمكن لنموذج التعلم العميق تمكين الآلات من فهم وتوليد اللغة البشرية [22]. تتضمن بعض التطبيقات الرئيسية للتعلم العميق في معالجة اللغة الطبيعية ما يلي:

- **التوليد التلقائي للنصوص:** يمكن لنموذج التعلم العميق تعلم مجموعة النصوص ويمكن إنشاء نص جديد مثل الملخصات والمقالات تلقائيًا باستخدام هذه النماذج المدربة.
- **ترجمة اللغة:** يمكن لنماذج التعلم العميق ترجمة النص من لغة إلى أخرى، مما يجعل من الممكن التواصل مع أشخاص من خلفيات لغوية مختلفة.
- **تحليل المشاعر:** يمكن لنماذج التعلم العميق تحليل مشاعر جزء من النص، مما يجعل من الممكن تحديد ما إذا كان النص إيجابيًا أو سلبيًا أو محايدًا. يتم استخدام هذا في تطبيقات مثل خدمة العملاء ومراقبة وسائل التواصل الاجتماعي والتحليل السياسي.

3.14.2 التعرف على الكلام: يمكن لنماذج التعلم العميق التعرف على الكلمات المنطوقة ونسخها، مما يجعل

من الممكن أداء مهام مثل تحويل الكلام إلى نص والبحث الصوتي والأجهزة التي يتم التحكم فيها بالصوت.

4.14.2 التعلم التعزيزي: في التعلم التعزيزي، يعمل التعلم العميق كعامل تدريب لاتخاذ إجراء في بيئة لتعظيم

المكافأة. تتضمن بعض التطبيقات الرئيسية للتعلم العميق في التعلم التعزيزي ما يلي:

برمجة الألعاب: تمكنت نماذج التعلم التعزيزي العميق من التغلب على الخبراء البشريين في ألعاب مثل Go و Atari Chess.

الروبوتات: يمكن استخدام نماذج التعلم التعزيزي العميق لتدريب الروبوتات على أداء مهام معقدة مثل الإمساك بالأشياء والملاحة والتلاعب.

أنظمة التحكم: يمكن استخدام نماذج التعلم التعزيزي العميق للتحكم في أنظمة معقدة مثل شبكات الطاقة وإدارة المرور وتحسين سلسلة التوريد.

يعتبر التعرف على المشي باستخدام بيانات مرئية (الصور والفيديو) أحد تطبيقات الرؤية الحاسوبية لذلك سوف نشرح بشكل مختصر الرؤية الحاسوبية وخصائصها وتطبيقاتها.

15.2 التعلم العميق والرؤية الحاسوبية

تعريف: الرؤية الحاسوبية (Computer Vision) هي مجال من مجالات الذكاء الاصطناعي (AI)

والتي تهدف إلى تمكين الأجهزة والأنظمة من فهم وتفسير واستخلاص المعلومات من العالم المرئي والذي يتضمن الصور الرقمية ومقاطع الفيديو وأي مدخلات مرئية أخرى.

تعمل الرؤية الحاسوبية بطريقة تحاكي الرؤية البشرية، باستثناء أن البشر لديهم السبق في مدة التعلم وكم

المعلومات البصرية. حيث يتمتع البصر البشري بميزة عمر السياق لتدريبه على كيفية تمييز الأشياء عن بعضها، و مدى بُعدها، وإذا كانت تتحرك أم ثابتة، وإذا كان هناك خطأ ما في الصورة المشاهدة.

1.15.2 تطور الرؤية الحاسوبية

قبل ظهور التعلم العميق، كانت المهام التي يمكن للرؤية الحاسوبية القيام بها محدودة للغاية وتتطلب

الكثير من الترميز اليدوي والجهد من جانب المطورين والمشغلين البشريين. على سبيل المثال يتطلب إجراء التعرف على المشي، تنفيذ الخطوات التالية:

إنشاء قاعدة بيانات: النقاط صور فردية لجميع الأشخاص المراد تتبعهم ضمن نفس التنسيق وهم يمشون.

التعليق على الصور: إدخال العديد من نقاط البيانات الرئيسية لكل صورة فردية، مثل المسافة بين الركبة والقدم وعرض القدم، والمسافة بين الأكتاف والحوض، وعشرات القياسات الأخرى التي تحدد الخصائص الفريدة لكل شخص.

النقاط صور جديدة: بعد ذلك، يجب النقاط صور جديدة، ومن ثم المرور بعملية القياس مرة أخرى، وتحديد النقاط الرئيسية في الصورة.

إجراء المقارنة: سيتمكن التطبيق بعد العمل اليدوي السابق من مقارنة القياسات في الصورة الجديدة بتلك المخزنة في قاعدة البيانات الخاصة به وتحديد ما إذا كانت تتوافق مع أي من الملفات التعريفية التي يتتبعها.

بناءً على ما سبق، كانت الأتمتة شبه غائبة، حيث تم تنفيذ معظم العمل يدويًا، مما جعل العملية عرضة للأخطاء وبمعدل دقة منخفض.

مع التعلم الآلي، لم يعد المطورون بحاجة إلى ترميز كل قاعدة يدويًا في تطبيقات الرؤية الخاصة بهم. بدلاً من ذلك، قاموا ببرمجة "الميزات"، وهي تطبيقات أصغر يمكنها اكتشاف أنماط معينة في الصور. ومن ثم استخدام أحد خوارزميات التعلم الإحصائي مثل الانحدار الخطي، والانحدار اللوجستي، وأشجار القرار أو آلات الدعم المتجه (SVM (Support Vector Machine لاكتشاف الأنماط وتصنيف الصور واكتشاف محتوياتها. ساعد التعلم الآلي في حل العديد من المشكلات التي كانت تشكل تحديًا تاريخيًا لأدوات وأساليب تطوير البرامج الكلاسيكية على سبيل المثال، فإن بناء ميزات البرنامج يتطلب جهود العشرات من المهندسين والاختصاصيين بالمشكلة المدروسة وهذا يستغرق وقتًا طويلاً للتطوير.

لقد قدم التعلم العميق نهجًا مختلفًا تمامًا للقيام بحل المشكلات اعتمادًا على الشبكات العصبونية، وهي وظيفة عامة الغرض يمكنها حل أي مشكلة يمكن تمثيلها من خلال الأمثلة. عندما تزود شبكة عصبونية بالعديد من الأمثلة المصنفة لنوع معين من البيانات، فستكون قادرة على استخراج الأنماط المشتركة بين تلك الأمثلة وتحويلها إلى معادلة رياضية تساعد في تصنيف المعلومات المستقبلية.

على سبيل المثال، يتطلب إنشاء تطبيق التعرف على المشي باستخدام التعلم العميق فقط تطوير أو اختيار خوارزميات مثل (CNN) وتدريبها بأمثلة لطريقة مشي الأشخاص الذين يجب اكتشافهم. مع وجود أمثلة كافية (الكثير من الأمثلة)، ستكون الشبكة العصبونية قادرة على التعرف على المشي دون تعليمات إضافية حول الميزات أو القياسات.

التعلم العميق هو طريقة فعالة للغاية للقيام بالرؤية الحاسوبية. في معظم الحالات، يتلخص إنشاء خوارزمية تعلم عميق جيدة في جمع كمية كبيرة من بيانات التدريب المصنفة وضبط المعلمات مثل نوع وعدد طبقات الشبكات العصبونية وعدد دورات التدريب.

2.15.2 طريقة عمل الرؤية الحاسوبية

تنقسم طريقة عمل الرؤية الحاسوبية إلى ثلاث مراحل أساسية:

الحصول على مجموعة البيانات: تطلب الرؤية الحاسوبية كميات كبيرة من البيانات. فهي تجري تحليلات للبيانات مرارًا وتكرارًا حتى تتمكن من تمييز الفروق والتعرف في النهاية على الصور والفيديو.

التدريب وبناء النموذج: يتم استخدام خوارزميات التعلم العميق (CNN للصور و RNN للفيديو).

لتدريب النموذج الافتراضي (model) وتعديل بارامتراته وذلك بتغذيته بمجموعة البيانات بحيث نحصل على أفضل بارامترات في نهاية التدريب، المودل الجيد يسمح للكمبيوتر بـ "النظر" للبيانات والتعرف عليها.

الاختبار: يتم اختبار النموذج السابق بتغذيته بصور أو فيديو جديد لتبيان مدى قدرة النموذج على التنبؤ.

3.15.2 أدوات برمجة الرؤية الحاسوبية

مكتبة OpenCV (Open Computer vision) هي مكتبة رؤية حاسوبية مفتوحة المصدر تعمل على معالجة الصور والفيديو.

مكتبة scikit-image (skimage) هي مكتبة رؤية حاسوبية مفتوحة المصدر تُستخدم لمهام معالجة الصور.

منصة TensorFlow عبارة عن منصة مفتوحة المصدر للتعلم الآلي والذكاء الاصطناعي منصة مفتوحة المصدر تقوم بمعالجة البيانات الصورية بطريقة سريعة جداً وفائقة الجودة.

مكتبة keras هي مكتبة تعلم عميق مفتوحة المصدر مبنية على TensorFlow .

الفصل الثالث

الأسس النظرية لخوارزميات وتقنيات التعلم العميق المستخدمة

Theoretical Foundations of Deep Learning

Algorithms and Techniques Used

الفصل الثالث

الأسس النظرية لخوارزميات وتقنيات التعلم العميق المستخدمة

1.3 مقدمة

أدى التقدم التكنولوجي السريع الذي يشهده عالمنا إلى تطوير العديد من خوارزميات وتقنيات التعلم العميق، يهدف هذا الفصل إلى تقديم شرح شامل ومبسط لأبرز هذه الخوارزميات، مع التركيز على بنيتها الداخلية وآلية عملها، تمهيداً لطريقة تطبيقها في الفصول القادمة من هذا البحث. علاوة على ذلك، يستعرض هذا الفصل أنواع البيانات الأساسية المستخدمة في أنظمة التعرف على المشي، مع تسليط الضوء على مجموعات البيانات التي سستخدم كمدخلات للشبكات العميقة المستخدمة.

2.3 الشبكة العصبونية التلافية

الشبكة العصبونية التلافية هي أحد أنواع الشبكات العصبونية التي تنتمي إلى التعلم العميق والمستخدم بشكل شائع في تحليل الصور تتألف بنيتها الأساسية من العديد من الطبقات التلافية وطبقات التجميع والطبقات المتصلة بشكل كامل حيث يمكنها تمثيل المعلومات المدخلة من خلال تسلسل هرمي من الميزات المحلية التي تستخرجها تلك الطبقات [23].

الشبكات العصبونية التلافية مناسبة بشكل خاص للعمل مع الصور نتيجة لاعتمادها المكاني القوي وثباتها بدرجة كبيرة بالنسبة للتغيرات [24]، فعلى سبيل المثال، بالنسبة لتصنيف الصور، يمكن للشبكات العصبونية التلافية استخراج سمات عميقة من البنية الداخلية للصورة ولذلك تعتبر أدوات فعالة لتجاوز هندسة السمات في مهام معالجة الإشارات (التعلم من البداية إلى النهاية) فهي تعتبر بمثابة مستخرجات تلقائية للميزات من الصورة وبالتالي تتجنب فقدان التفاعل المكاني الذي تسببه الطرائق التقليدية للتعلم الآلي، ومع ذلك، تتطلب الشبكات العصبونية التلافية، مئات الأمثلة في كل فئة للتعلم الفعال.

1.2.3 ميزات الشبكة العصبونية التلافية

- تقليل حجم البيانات المدخلة: تعتمد الشبكات العصبونية التلافية (CNNs) على معلومات وحدات البكسل المجاورة لمعالجة الصور بشكل فعال. تقوم في البداية بتقليل حجم البيانات باستخدام الطبقات التلافية (Convolutional Layers)، ثم تمرر النتائج إلى طبقة التنبؤ (Prediction Layer) في النهاية لتحليل البيانات واتخاذ القرارات [25].
- تجمع بنية CNN ضمناً بين الفوائد التي تم الحصول عليها من تدريب الشبكة العصبونية القياسية وعملية التلاف للتعامل بكفاءة مع المهام المطلوبة مثل التصنيف والتعرف والتحديد.
- التعامل مع مجموعات البيانات الكبيرة: تعد CNN أداة مثالية للعمل مع مجموعات البيانات الكبيرة، حيث تستفيد من وفرة البيانات لتحسين أدائها وزيادة دقتها في التعرف والتصنيف.

2.2.3 تحديث الأوزان في الشبكة العصبونية التلافية

يتكون نظام التعرف بإشراف من أربع مكونات أساسية كما هو موضح بالشكل (1-3):

- طبقة الدخل (X) والتي تكون عبارة عن مجموعة بيانات (dataset) معنونة.
 - النموذج: ويتكون من مجموعة من الطبقات ومهمته الحصول على علاقة بين الدخل والخرج المتوقع ($\hat{y}$)
- $$\hat{y} = h(X)$$
- دالة الخطأ: هي الطريقة التي ستمكن بها الشبكة من قياس مدى جودة العمل الذي تقوم به بالنسبة لبيانات التدريب الخاصة بها، وبالتالي كيف ستمكن من توجيه نفسها في الاتجاه الصحيح عن طريقها تحدد إشارة التغذية الراجعة المستخدمة للتعلم وتعطى بالعلاقة (1-3) [26]:

$$J(w) = \frac{1}{2m} \sum_{i=0}^m (y_i - \hat{y}_i)^2 \quad (1-3)$$

حيث i هو رقم العينة و y_i هي القيمة الحقيقية لفئة العينة i .

- المحسن (optimizer): وهو خوارزمية تقوم باختيار أفضل البارامترات للنموذج وهي الأوزان (W) والانحيازات

(b) حيث ستقوم الشبكة من خلال هذه الخوارزمية بتحديث نفسها بناءً على البيانات التي تراها ودالة الخسارة

الخاصة بها.

1.2.2.3 خوارزمية الانتشار الخلفي (Back Propagation) هي الخوارزمية التي من خلالها يتم تحديد

قيم الأوزان الجديدة وسنشرح الحاجة لها وطريقة تطبيقها في هذه الفقرة. لنفترض أن لدينا مجموعة بيانات

التدريب X على الشكل التالي:

$$X = \begin{bmatrix} x^{1t} \\ x^{2t} \\ x^{3t} \\ \vdots \\ x^{mt} \end{bmatrix}$$

حيث أن $x^{1t}, \dots, x^{mt}$ هي عينات الدخل بعد إجراء عملية المنقول عليها.

ومجموعة الخرج الحقيقية هي:

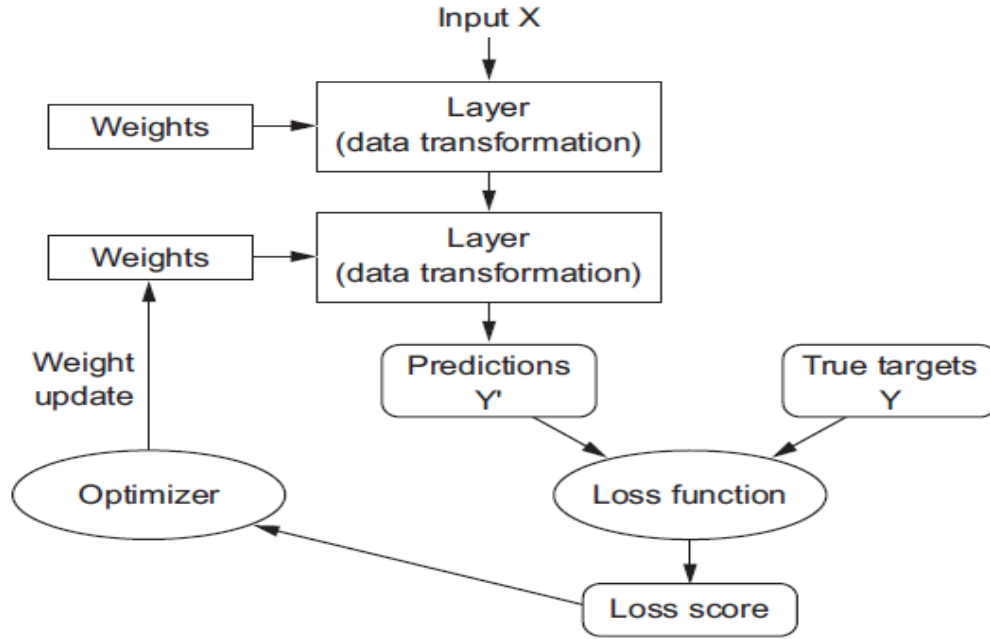
$$Y = \begin{bmatrix} y^1 \\ y^2 \\ y^3 \\ \vdots \\ y^m \end{bmatrix}$$

وبما أن علاقة الدخل بالخرج تعطى بالعلاقة (2-3) [27]:

$$h_w(x^i) = (x^i)^T \cdot W \quad (2 - 3)$$

حيث أن $h_w(x^i)$: دالة الخرج التي تطبق الأوزان W على متجه الإدخال X .

نستطيع أن نقول:



الشكل (1-3) المخطط الصندوقي لنظام التعرف [26]

$$X.W - Y =$$

$$\begin{pmatrix} (x^1)^T \cdot W \\ \vdots \\ (x^i)^T \cdot W \end{pmatrix} - \begin{pmatrix} y^1 \\ \vdots \\ y^m \end{pmatrix} = \begin{pmatrix} h_w(x^1) - y^1 \\ \vdots \\ h_w(x^i) - y^m \end{pmatrix}$$

وباستخدام القانون الرياضي الذي يقول (3-3) [28]

$$z^T \cdot z = \sum_i (z_i)^2 \quad (3-3)$$

سيكون لدينا

$$\frac{1}{2} (X.W - y)^T (X.W - y) = \frac{1}{2} \sum_{i=1}^m (h_w(x^i) - y^i)^2 = J(W)$$

أي أن الناتج سيساوي علاقة الخطأ المحددة بالعلاقة (1-3) .

باشتقاق معادلة الخطأ السابقة ومساواتها بالصفر ينتج لدينا معادلة التسوية (normal equation) التي تعطى

بالعلاقة (4-3) [27]:

$$W = (X^T X)^{-1} X^T Y \quad (4-3)$$

من الصعب جداً إيجاد حلول لمعادلة التسوية السابقة وذلك للأسباب التالية:

- الحل الحسابي بطيء جداً (X عبارة عن مصفوفة كبيرة، وحساب عكس المصفوفة سيكون مكلف وصعب جداً).
- يستخدم جميع نقاط البيانات وبالتالي قد يكون مفرطاً في التدريب.
- لا يعمل هذا الحل للتصنيف وإنما فقط للانحدار نظراً لأن بعض المعادلات ستختفي وسيكون لدينا حلول لا نهائية (يمكن للخطوط اللانهائية أن تفصل الفئات عندما يكون متوسط الخطأ التربيعي يساوي الصفر).

تم اللجوء إلى خوارزمية الانتشار الخلفي (Back Propagation)، في حالة الخسارة، فإن التدرج $\frac{\delta \text{loss}}{\delta w}$

هو مقدار تغير الخسارة عند تغير أحد الأوزان بمقدار صغير. إنه يوجه بحثنا عن الأوزان التي تقلل الخسارة؛ إذا غيرنا w قليلاً، وتأثرت الخسارة كثيراً، فهذا يعني أن الخسارة حساسة لهذا التغيير.

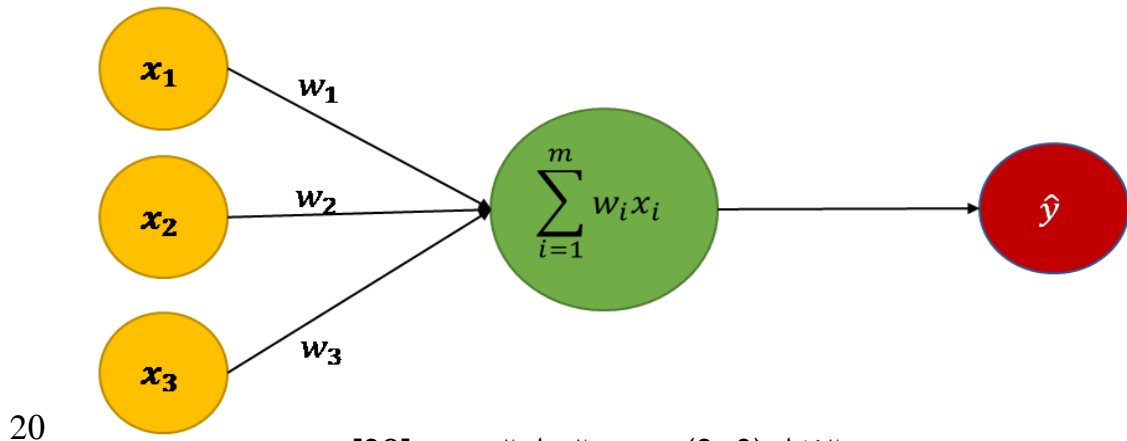
نظراً لأن الخسارة محدبة دائماً، فهذا يعني أن أي تغيير في w يحرك الخسارة بعيداً عن الحد الأدنى. لذا إذا تحركنا عكس هذا التغيير، فإننا نصل إلى الحد الأدنى. إذا كانت قيمة التغيير هذه صغيرة، فهذا يعني أننا بحاجة إلى التحرك خطوة صغيرة، والعكس صحيح. إذا كان لدينا طبقة واحدة من الخلايا العصبونية/السمات، يمكننا بسهولة معرفة حساسية الخسارة للعصبون $\frac{\delta \text{loss}}{\delta w_i}$ ، حيث w_i هو وزن العصبون للخلية العصبونية/السمات i إلى المخرجات إذا كان تابع التفعيل هو الوحدة الخطية المقومة (ReLU) فإن الخرج يعطى بعلاقة الضرب النقطي للأوزان مع دخل الخلية العصبونية مضافاً إليها الانحياز ويمكن التعبير عن معادلة الخرج بالعلاقة (5-3) [29].

$$\text{output} = \text{ReLU}(\text{dot}(W, \text{input}) + b) \quad (5-3)$$

يحتوي النموذج على معلمات وهي الأوزان W والانحيازات b . وتسمى هذه المعلمات بالمعلمات القابلة للتدريب لأن قيمها تتغير أثناء التدريب، تحتوي هذه الأوزان على المعلومات التي تعلمتها الشبكة من بيانات

التدريب. في البداية، يتم ملء مصفوفات الأوزان بقيم عشوائية صغيرة. وفي نهاية التدريب يصبح الخرج مساوياً لـ $\text{ReLU}(\text{dot}(W, \text{input}) + b)$ حيث يتم إيجاد ما يسمى مجموع الدخل الموزون (weighted sum) المساوي لقيم مجاميع كل دخل بالوزن المقابل له كما هو مبين بالشكل (2-3).

إن عملية التعلم تأتي بعد تهيئة المعلمات بقيم عشوائية حيث تضبط الأوزان تدريجياً، باستخدام التغذية الراجعة المتتالية. هذا التعديل التدريجي، والذي يُطلق عليه أيضاً التدريب، هو في الأساس التعلم الذي يدور حوله التعلم الآلي. إن الغاية من التدريب هو تصغير الخطأ إلى أصغر قيمة ممكنة وبالتالي أثناء



الشكل (2-3) مجموع الدخل الموزون [29]

التدريب في كل دورة يتم حساب قيمة المشتق وإضافته إلى الوزن حتى نحصل على الوزن الأقرب للمثالي.

لي وذلك باستخدام الخطوات التالية:

- تحديد مجموعة من عينات التدريب x والأهداف المقابلة y .
- القيام بتدريب الشبكة على x (المسار الأمامي) للحصول على التنبؤات y_{pred} .
- حساب خسارة الشبكة في الدفعة، وهو مقياس لعدم التطابق بين y و y_{pred} .
- القيام بتحديث جميع أوزان الشبكة بطريقة تقلل قليلاً من الخسارة (المسار الخلفي) في هذه الدورة.

إن تحديث أوزان الشبكة يعتمد على تطبيق خوارزمية المسار الخلفي (Backpropagation) وذلك من

خلال حساب المشتقات وتحديث المعلومات. من الشكل (3-3) الذي يمثل مداخل ومخرج أحد العصبونات أثناء

تحديث الأوزان يكون مشتق الخطأ بالنسبة للخروج z يساوي $\frac{\partial l}{\partial z}$ وهي تمثل مساهمة z بالخطأ.

كي نستطيع إدخال هذا الخطأ في المرحلة السابقة نوجد تغير الطبقة السابقة z بالنسبة ل x ويساوي

$$\frac{\partial z}{\partial x} \text{ وبالتالي إذا ضربنا بالمقدار } \frac{\partial l}{\partial z} \text{ نحسب مساهمة } x \text{ بقيمة الخطأ سينتج } \frac{\partial z}{\partial x} * \frac{\partial l}{\partial z} \text{ والذي}$$

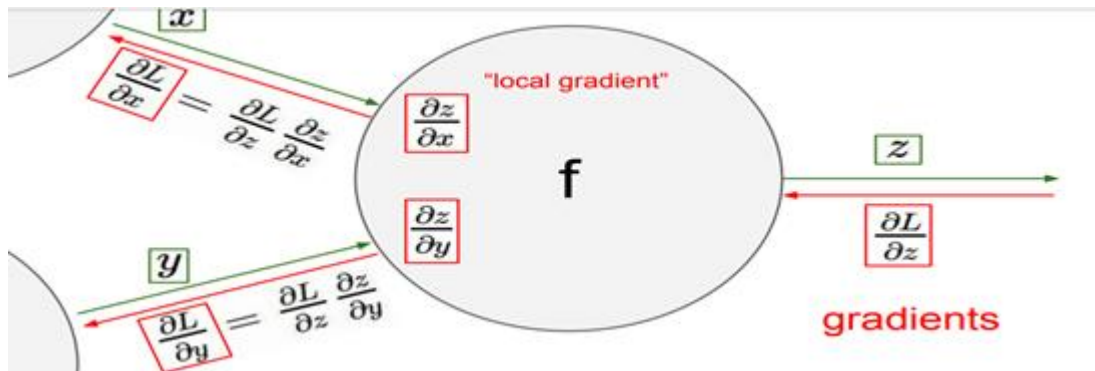
يساوي $\frac{\partial l}{\partial x}$ وهو مقدار الخطأ الذي يسببه x وبالتالي عند تحديث الأوزان يعطى الوزن الجديد $W(\text{new})$ بالعلاقة

(6-3) [30] التالية:

$$W(\text{new}) = W(\text{old}) + \delta W \quad (6-3)$$

$$\delta W = -\mu \frac{\partial j(w)}{\partial w} \text{ حيث } W = W(\text{old}) \text{ وهو وزن العصبون قبل التحديث}$$

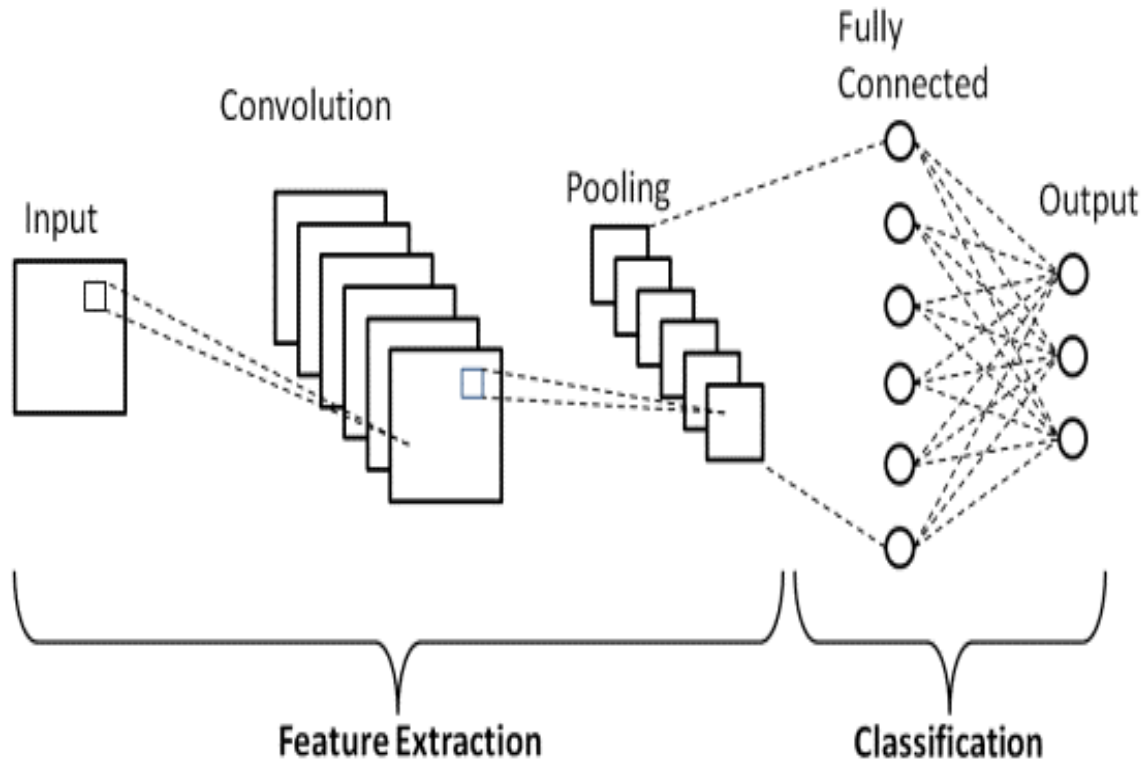
تعتبر μ عن سرعة الخطوة باتجاه القيمة الصغرى للخطأ والتي يمكن أن نسميها سرعة التعلم W_{new} هو قيمة الوزن.



الشكل (3-3) مداخل ومخرج عصبون في الشبكة العصبونية التلافيفية أثناء تحديث الأوزان [30]

3.2.3 طبقات الشبكة العصبونية التلافيفية

يبين الشكل (3-4) بنية الشبكة العصبونية التلافيفية حيث تتألف من الطبقات التالية:



الشكل (3-4) بنية الشبكة العصبونية التلافية [30]

• طبقة الإدخال Input: تتكون طبقة الادخال من البيانات الخام للصورة اللازمة لتغذية الشبكة العميقة.

• طبقة التلاف Convolution layer .

• طبقة الانتقاء Pooling Layer.

• الطبقة التامة الربط Fully Connected Layer (FC) .

• الإخراج Output.

1.3.2.3 طبقة التلاف (Convolutional layer)

تستخدم لاستخراج الميزات من مجموعة البيانات المدخلة. وهي تطبق مجموعة من المرشحات القابلة للتعلم

المعروفة باسم النواة على الصور المدخلة. المرشحات عبارة عن مصفوفات أصغر حجمًا باستخدام مصفوفتي

الصورة والمرشح، نجري عملية التلاف من خلال تطبيق حاصل الضرب النقطي [30] من الزاوية العلوية اليسرى

للصورة إلى اليمين. ثم تجمع قيم حاصل الضرب، تتوافق القيمة الناتجة مع القيمة الأولى (الزاوية العلوية اليسرى) في المصفوفة الجديدة. تتحرك النواة لأسفل بمقدار قيمة الخطوة التمريرية (stride)، ينزلق المرشح فوق بيانات الصورة المدخلة حتى يسمح الصورة بالكامل ويعطي صورة جديدة بأبعاد أقل، هذا البعد يعطى بالعلاقة

(7-3):[31]

$$M_2 = \frac{M_1 - F + 2P}{s} + 1 \quad (7-3)$$

حيث أن M_2 هو قيمة بعد الصورة الناتجة و M_1 هو قيمة أبعاد الصورة الأساسية S ، هو قيمة الخطوة التمريرية و P هو التبطين الصفري (padding) وهو عملية إضافة أصفار حول الصورة لتعويض النقص بسبب عملية التلاف تسمى الصورة الناتجة خريطة الميزات لأنها تحوي على الميزات المستخرجة من الصورة إن البعد الثالث للصورة الناتجة والذي يسمى العمق (depth) يساوي عدد المرشحات المستخدمة في كل عملية التلاف كما هو موضح بالشكل (3-5).

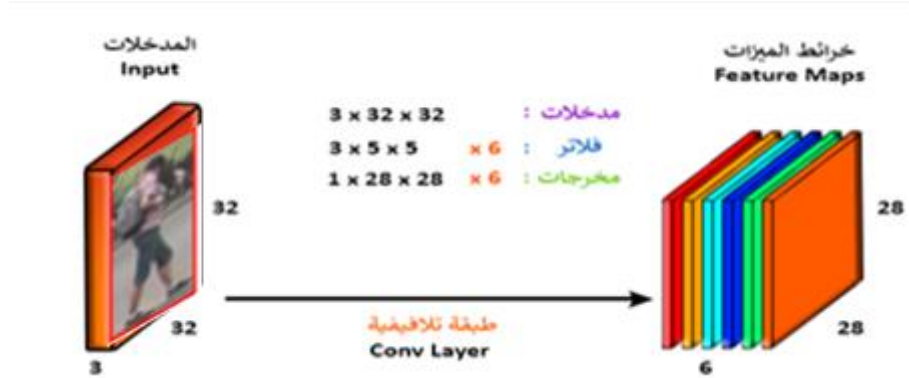
بعد ذلك نقوم بتطبيق دالة التنشيط على النتائج (activation function) والتي تعتبر جسر العبور إلى الطبقة التالية في الشبكة العصبونية التلافية وتكمن فائدتها في تقليل كمية الحسابات المنجزة من خلال عدم تفعيل جميع نقاط خرائط الميزات في آن واحد، وبمعنى أدق تفعيل النقاط التي تمثل الميزات وإقصاء النقاط التي لا تمثلها.

ناتج الطبقة التلافية نسميه أوزان (weights) وهي تتعدل عند كل طبقة تلافية داخل النموذج العميق المستخدم وحتى انتهاء عملية التدريب.

2.3.2.3 طبقة الانتقاء (pooling layer)

تهدف هذه الطبقة إلى تقليل أبعاد خريطة الميزات المصححة واختزالها (down sampling) واستخلاص أهم المعلومات في الصور (الميزات) [32] وأكثر أنواع التجميع انتشارا هو التجميع الأقصى (maxPooling) حيث

تؤخذ القيمة الكبرى بين القيم والتجميع المتوسط (Average Pooling) حيث تؤخذ القيمة المتوسطة للقيم التي يطبق عليها التجميع.



الشكل (3-5) تطبيق الالتفاف على الصورة [32]

أهم بارامترات هذه الطبقة والتي تحتاج الى تعيين هي حجم نواة التجميع وطول خطوة التجميع.

3.3.2.3 الطبقة تامة الربط (fully connected layer)

يكون الإدخال الى العقدة عبارة عن قيمة خطية من مخرجات الطبقة السابقة وعبر هذه الطبقة يتم

اختيار دالة التنشيط المناسبة للمهمة المطلوبة من نموذج التعلم العميق التي نريد حلها. أهم بارامترات هذه الطبقة هي تابع التفعيل وعدد الوحدات المناسب لمهمة التصنيف.

4.3.2.3 طبقة الإخراج: وهي في هذا البحث طبقة التعرف regression layer وتحتوي على هذه البارامترات:

- تابع الخسارة (loss function): هو جزء أساسي من أي نموذج تعلم آلي فهو يحدد هدفًا (التعرف على سبيل

المثال) يتم قياس أداء النموذج على أساسه.

- معدل التعلم (learning rate): تساعد على تقارب النموذج بشكل أسرع . إن اختيار معدل تعلم خاطئ، سواء

كان صغيرًا جدًا أو كبيرًا جدًا، يمكن أن يكون له تأثير كبير على الناتج.

- المحسّن (optimizer): يستخدم لتعديل الأوزان بعد حساب قيمة الخطأ.

4.2.3 استخراج الميزات

بالنسبة للشبكة العصبونية التلافية، فإنها تكتشف تلقائيًا السمات المهمة [33] مثل الرأس والأذنين والساقين واليدين والجذع وما إلى ذلك، تعمل الطبقات المتصلة بالكامل بعد ذلك كمصنف أعلى لهذه السمات.

تعد طبقات التلاف (Convolutional Layers) العنصر الأساسي والمحوري في الشبكات العصبونية الالتفافية (CNN) ، حيث تُتيح للنموذج اكتشاف السمات المهمة تلقائيًا من البيانات. تعمل هذه الطبقات بطريقة هرمية؛ تبدأ باستخراج السمات البسيطة مثل الحواف والزوايا، ثم تجمع الطبقات التالية هذه السمات لتحديد الأشكال البسيطة مثل الخطوط أو الأقواس. بعد ذلك، تُدمج السمات الأكثر تعقيدًا في الطبقات الأعمق لاستخلاص مكونات أكثر تميزًا مثل أجزاء الجسم أو الأجسام الكاملة.

تجدر الإشارة إلى أن النموذج لا يعرف مسبقًا ماهية هذه العناصر، مثل الوجه أو اليدين، بل يتعلم التعرف عليها من خلال الأنماط المتكررة التي يراها في الصور أثناء التدريب. في النهاية، تستخدم الطبقات المتصلة هذه السمات المستخلصة من مرشحات التلاف لتصنيف الصور بدقة بناءً على الأنماط المكتشفة.

5.3 محول الرؤية (ViT) Vision Transformer

لقد قدمت المحولات في عام 2017 بواسطة Vaswani et al في الورقة البحثية الشهيرة "attention is all you need". الوظيفة الأساسية التي تدعم هذه النماذج هي آلية الانتباه (attention) [34].

تطبق نماذج المحولات (Transformers) آلية الانتباه على مستويات متعددة من الطبقات، مما يمنحها القدرة على التعامل مع العلاقات المعقدة بين عناصر البيانات بكفاءة عالية. تتكون المحولات من مجموعة من الكتل التشفيرية (Encoder Blocks) ، والتي تُمثل الوحدات الأساسية لهذه النماذج.

كتل المشفر تشفر المزيد من المعلومات في تضمينات الرمز (token) باستخدام سياقها. تنتج هذه العملية تمثيلاً دلاليًا أعمق لكل رمز، في نهاية هذه العملية، نحصل على تضمينات (embedding) غنية بالمعلومات. هذه التضمينات هي الناتج النهائي لجوهر المحول، بما في ذلك محول الرؤية (vision transformer) المسمى اختصاراً ViT.

يتم استخدام مجموعة أخرى من الطبقات لتحويل هذه التضمينات الغنية إلى تنبؤات مفيدة، تسمى هذه الطبقات القليلة الأخيرة "الرأس" (head) ويتم استخدام رأس مختلف لكل مهمة، مثل التصنيف، والتعرف على الكيانات والإجابة على الأسئلة. تعمل المحولات المرئية بنفس الطريقة ولكن بدل من تقسيم الكلام إلى رموز فهي تقسم الصور إلى رقع (patch).

إن عمل محول الرؤية (ViT) الموضحة بنيته بالشكل (3-6) يتم من خلال تنفيذ المراحل التالية:

1- تضمين الدفعات patch embeddings: من أجل تغذية الصور في نموذج ViT (Vision transformer)، يتم تقسيمها أولاً إلى شبكة من الدفعات (رقع) غير المتداخلة، حيث يتم التعامل مع كل منها كرمز منفصل. ثم يتم تسطيح هذه الرقع في تسلسل أحادي الأبعاد وتميريرها عبر طبقة خطية لتحويلها إلى تسلسل من التضمينات. بافتراض أن عدد الرقع هو N . يتم تحقيق هذا الإسقاط الخطي بواسطة CNN بحجم نواة $K_H * K_W$ (H يمثل الارتفاع و W يمثل العرض) يتم الإشارة إلى كل رقعة patch باسم P_i حيث أن i هي رقم فهرس الدفعة بينما يشار إلى التضمين على أنه $E(P_i)$ [34].

2- تضمين الموقع Positional Embeddings: نظرًا لأن نموذج ViT لا يستخدم الطبقات التلافية مثل نماذج تصنيف الصور التقليدية، فإنه يحتاج إلى وسيلة لدمج المعلومات المكانية في تمثيلاتها للحفاظ على معلومات الموضع الخاصة بالرقع، يتضمن ViT تضمينات موضعية. تساعد التضمينات الموضعية النموذج على فهم العلاقات المكانية بين الرقع المختلفة، مما يمكنه من التقاط بنية الصورة حيث يضاف الموضع إلى التضمينات السابقة [35].

3- مُشَفِّر المحول Transformer Encoder: إن جوهر نموذج ViT هو مجموعة من طبقات مُشَفِّر المحول، والتي تعالج تسلسل التضمينات الموضعية. تتكون كل طبقة مُشَفِّر محول في نموذج ViT من طبقتين فرعيتين: آلية الانتباه الذاتي متعدد الرؤوس (multi-head self-attention) وشبكة تغذية أمامية حسب الموضع (Multi-Head Position-wise feedforward network).

- **آلية الانتباه الذاتي متعدد الرؤوس:** هذه الطبقة الفرعية في نماذج المحولات مسؤولة عن معالجة تضمينات الإدخال لالنقاط المعلومات السياقية الأكثر أهمية. تعتمد العملية على آلية الانتباه الذاتي (Self-Attention)، حيث يقوم كل تضمين إدخال "بالاهتمام" بجميع التضمينات الأخرى في التسلسل ذاته، لتحديد مدى أهميتها بالنسبة له من خلال حساب التشابه بينها. بناءً على هذا التشابه، يتم تخصيص أوزان تعكس مدى ارتباط التضمينات ببعضها البعض، ثم يتم دمج هذه التضمينات باستخدام مجموع الأوزان (Weighted Sum) للحصول على تمثيل جديد يعكس السياق الكلي.

تعمل آلية الانتباه الذاتي بطريقة متعددة الرؤوس (Multi-Head Attention)، حيث يتم تنفيذها بشكل متوازٍ عبر عدة رؤوس، لكل منها مجموعة مستقلة من الأوزان. هذا التوازي يسمح لكل رأس بتعلم تمثيل مختلف لتسلسل الإدخال، ما يُمكن النموذج من النقاط أنواع متنوعة من التبعية والعلاقات داخل البيانات. في النهاية، يتم دمج هذه التمثيلات لتقديم فهم شامل للعلاقات السياقية.

- **شبكة تغذية أمامية متعددة الطبقات أو شبكة تغذية أمامية حسب الموضع:** يتم تمرير تضمينات الإدخال عبر شبكة تغذية أمامية حسب الموضع، تطبق هذه الطبقة الفرعية شبكة عصبية أمامية متصلة بالكامل على كل موضع في التسلسل بشكل مستقل. تتكون هذه الشبكة من تحويلين خطيين، مفصولين بواسطة دالة تنشيط غير خطية (عادةً ReLU)، ويتم تطبيقها على كل موضع في التسلسل بشكل مستقل.

الإخراج النهائي لـ Transformer Encoder هو العلامات (tokens) التي يتم تقسيمها إلى علامات

أساسية أو (يُشار إليه بـ F) وعلامات محلية (يُشار إليها بـ Q). يتم نقل هذه الرموز إلى مصفوفة ثنائية الأبعاد، حيث يكون بُعد هذه المصفوفة هو $M = \sqrt{N}$ حيث N هو عدد الرقع بافتراض أن الأبعاد الرئيسية للإدخال هي H و W ، فيجب اختيار N وفقاً للمعادلة (9-3) [36] (S هي الخطوة الأساسية):

$$N = \left(\frac{H-K_H}{S} + 1 \right) * \left(\frac{W-K_W}{S} + 1 \right) \quad (9-3)$$

بعد ذلك، يتم إضافة تضمين الموضع $(E(x_i^l))$ إلى تضمين الفئة القابلة للتعلم، حيث يمثل l موقع الدفعة، الذي تحتفظ حالة خرجها بمعلومات الصورة بأكملها وتعمل كمتجه أساسي. ثم تتم إضافة المتجهات الناتجة مع تضمينات الموضع l للحفاظ على معلومات الموضع. بعد ذلك، يتم تغذية التسلسل النهائي للمتجهات كحالة ابتدائية z_0 في مُرمِّز المحول لتوليد متجهات ميزة $N + 1$ حيث N هو عدد الرقع بالإضافة إلى تضمين الفئة. تعطى قيمة الحالة الابتدائية z_0 التي تحوي على تسلسل الفئة وتضمينات كل رمز في الإدخال بالعلاقة (10-3) [37]:

$$z_0 = [X_{class}, E(x_1^l), E(x_2^l), \dots, E(x_N^l)] + l \quad (10 - 3)$$

حيث أن X_{class} هي الفئة الرئيسية القابلة للتعلم.

يتكون مُشغِّل المحول من إجمالي كتل B . تحتوي كل كتلة على كتل الانتباه متعدد الرؤوس MSA

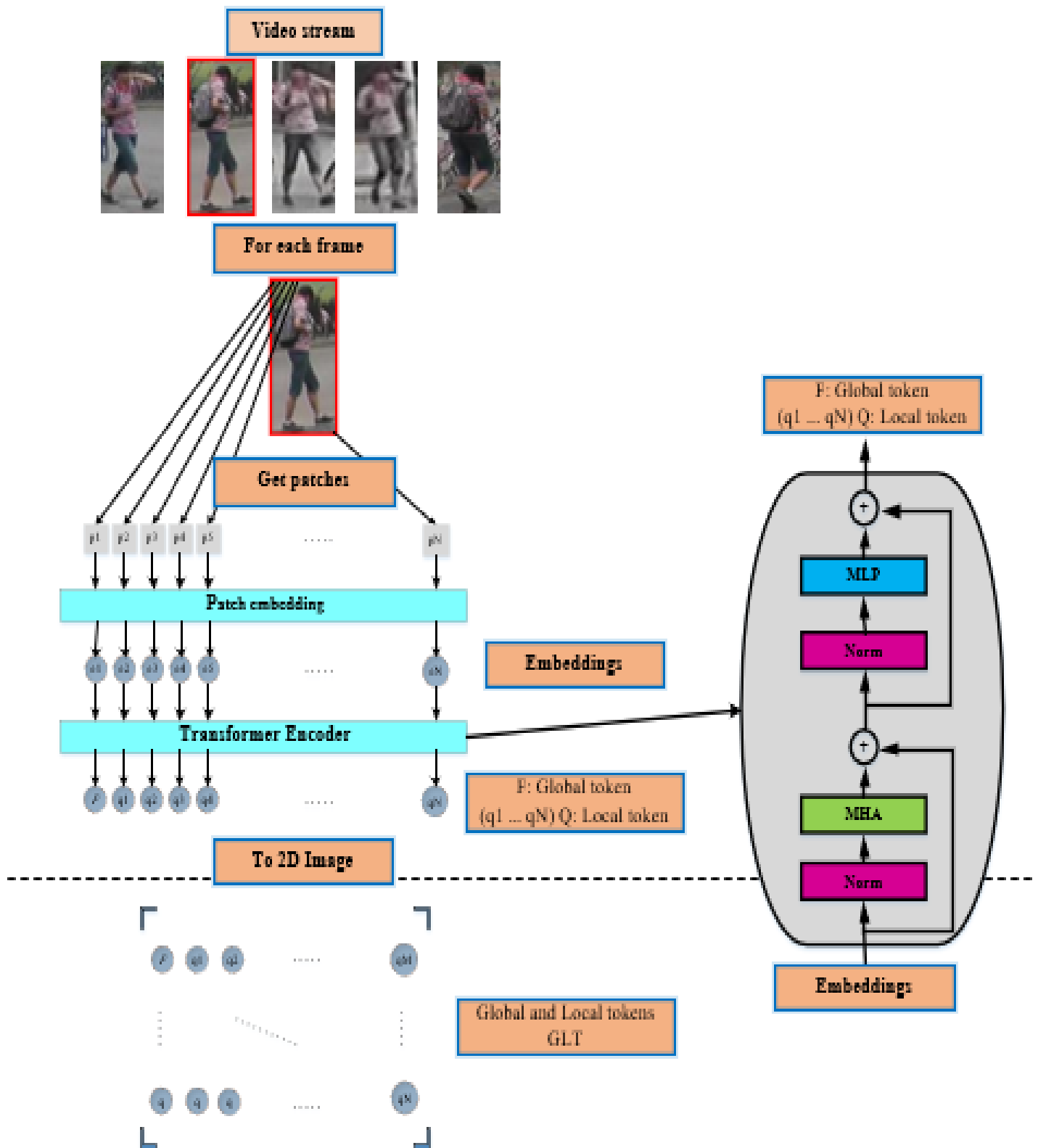
وكتل التغذية الأمامية MLP متناوبة. يتم تطبيق طبقات استنظام (LN) Layer norm قبل كتل MSA و MLP

و بعد كل مُشغِّل لضمان استقرار القيم المحسوبة وتحسين كفاءة عملية التدريب ، يمر خرج مُشغِّل المحول F

الموضح في المعادلة (11-3) [38] عبر جميع كتل B حيث أن z_B هو خرج الطبقة B الذي يعطى بالعلاقة

$$(12-3)(38).$$

$$F = \text{LN}(z_B) \quad (11 - 3)$$



الشكل (3-6) البنية العامة للمحول المرئي [37]

$$z_B = z_{B-1} + \text{MSA}(\text{LN}(z_{B-1})) + \text{MLB}(z_{B-1} + \text{MSA}(\text{LN}(z_{B-1}))) \quad (12 - 3)$$

بفرض أن $N = 100$ ، ثم $M = 10$ ، ثم يتم حساب متوسط هذه العلامات للحصول على علامات محلية باستخدام طبقة التجميع، ويحتوي كل صف على M رمز ويمثل كل صف متوسط الرموز المحلية و العامة G (Global Local Token) ثم يتم تصنيفها باستخدام مصنف كما هو موضح بالشكل (3-7) يتم تحديث خرج المحول بالنسبة للرموز العامة بالعلاقة (3-13) [38] والرموز المحلية بالعلاقة (3-14) [38].

$$F' = F + \varepsilon G \quad (13 - 3)$$

$$q_i = q_i + \varepsilon G \quad (14 - 3)$$

G هي العلامة الأساسية ، ε هي معامل صغير عشوائي.

L هي متوسط GLT الذي تم الحصول عليه بعد تجميع متوسط Q و G ويعطى بالعلاقة (3-15) [38]

$$L_u = \frac{1}{M} \sum_{j=u*M+1}^{(u+1)*M} \frac{Q_j + \varepsilon G}{1 + \varepsilon G} \quad u = 1, \dots, M \quad (15-3)$$

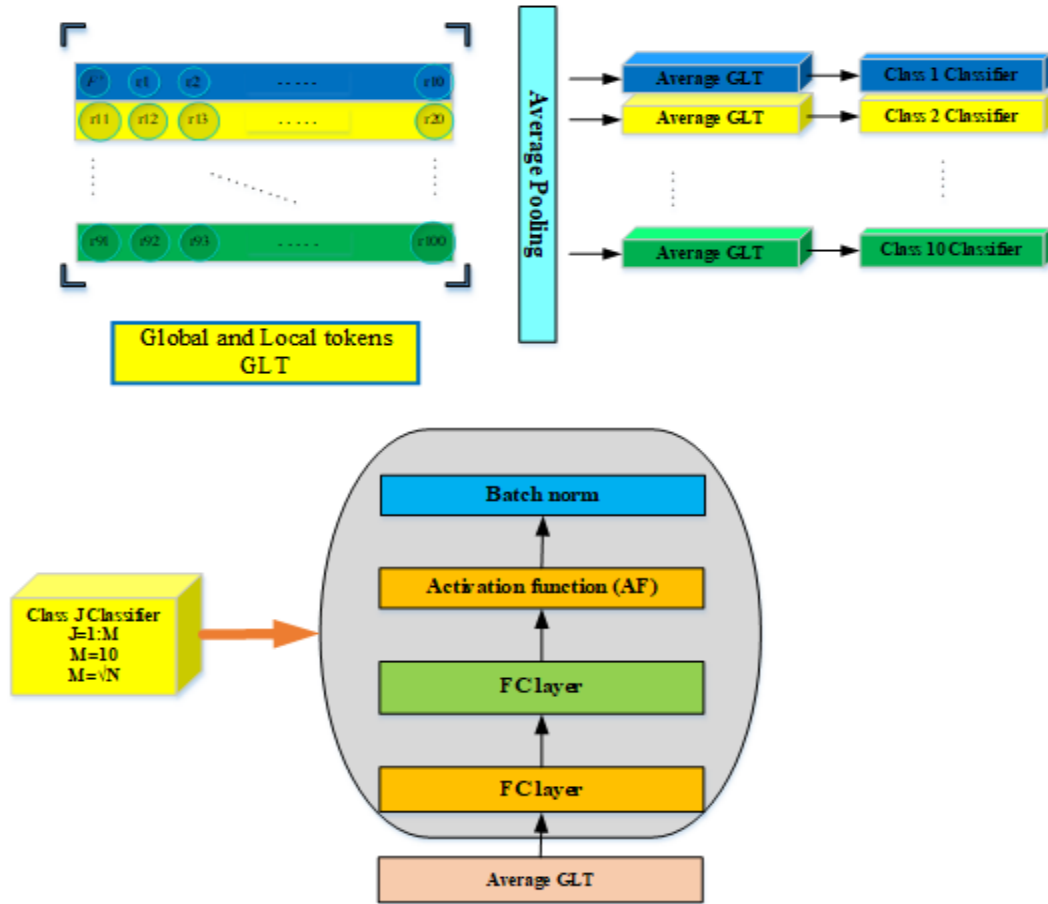
حيث M هو عدد الرقع في الصورة.

المصنف: يوجد في الجزء العلوي من نموذج ViT رأس تصنيف يأخذ مخرجات طبقة ترميز المحول الأخيرة ويقوم بربطها باحتمالات الفئة باستخدام طبقة متصلة بالكامل ووظيفة التنشيط باستخدام Softmax.

الخرج النهائي للمحول هو خرج الطبقة المتصلة بالكامل (FC) ويعطى بالعلاقة (3-16) [38]:

$$y_u = FC_u(L_u) \quad u = 1, \dots, M \quad (16 - 3)$$

يتم تمرير المخرجات y من خلال دالة التنشيط (Activation Function) AF ويتم جمع النتائج معًا. يمثل معامل



الشكل (7-3) مصنف المحول [39]

الحد الأقصى للنتيجة معرف الشخص (ID) حيث يتم حساب الميز بالعلاقة (17-3) [38]:

$$\text{score} = \sum_{u=0}^M \text{AF}(y_u) \quad (17-3)$$

تعطى قيمة التوقع prediction بالعلاقة (18-3) [38] حيث يحدد التابع argmax القيمة الكبرى للميز.

$$\text{prediction} = \text{argmax}(\text{score}) \quad (18-3)$$

6.3 الشبكات العصبونية التكرارية (Recurrent Neural Network)

الشبكات العصبونية التكرارية هي نوع من الشبكات العصبونية التي تتميز بوجود حلقات تغذية مرتدة داخل

بنيتها، مما يسمح بتدفق المعلومات ذهاباً وإياباً بين طبقات الشبكة. هذا التصميم يمكّنها من التعامل مع البيانات

المتسلسلة وزمنية الطبيعة، مثل التعرف على الكلام والترجمة الآلية. تتشابه بنية الشبكة العصبونية التكرارية مع

الشبكات العصبونية ذات التغذية الأمامية، حيث تتكون من مجموعة من الطبقات، لكن الفرق الجوهرى هو أن الطبقات في الشبكة التكرارية متصلة في حلقة، مما يُتيح لمخرجات كل طبقة أن تُعاد تغذيتها كمدخلات لنفس الطبقة. يوضح الشكل (3-8) هذا الفرق بين الشبكات ذات التغذية الأمامية والشبكات التكرارية.

تمتلك الشبكات العصبونية التكرارية ما يُعرف بـ "الذاكرة"، حيث يمكنها استخدام المعلومات المستخلصة من المدخلات السابقة للتأثير على المعالجة الحالية. على عكس الشبكات العصبونية العميقة التقليدية التي تفترض استقلالية المدخلات والمخرجات، تعتمد مخرجات الشبكات التكرارية على السياق المتتابع للعناصر داخل التسلسل، تستخدم الشبكات العصبونية التكرارية لحل المشكلات الزمنية أو الترتيبية، مثل معالجة اللغة الطبيعية والتعرف على الكلام وترجمة الصور، وذلك لأنها قادرة على فهم العلاقات الزمنية أو السياقية بين نقاط البيانات، مما يجعلها أساسية في العديد من التطبيقات التي تعتمد على تحليل التسلسلات الزمنية.

تتكون الشبكة العصبونية التكرارية من سلسلة من الوحدات المتكررة، حيث تمثل كل وحدة خطوة زمنية معينة. تحتوي كل وحدة على حالة داخلية تُعرف باسم الحالة المخفية (Hidden State)، وهي تمثل المعرفة التي تحتفظ بها الشبكة في تلك اللحظة الزمنية، بناءً على المعلومات السابقة التي مرت عبر الشبكة.

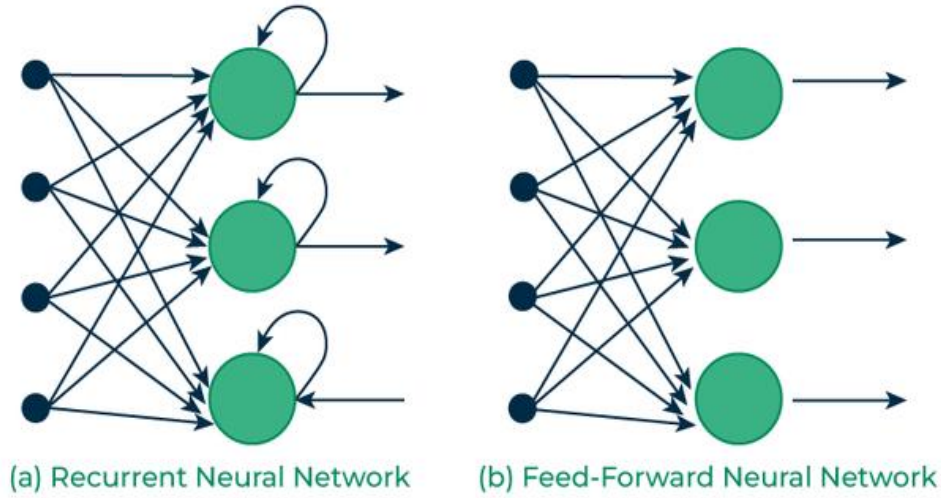
عند فك الحلقات في الشبكة التكرارية (Unrolling)، تصبح الشبكة عبارة عن سلسلة من الوحدات المتصلة، حيث يُظهر الشكل (3-9) كيف يتم تمثيل البنية الزمنية للشبكة عند فكها. هذا الشكل يوضح كيفية تمرير الحالة المخفية من خطوة زمنية إلى أخرى. تتم عملية تحديث الحالة المخفية في كل خطوة زمنية لتعكس التغير في المعرفة بناءً على البيانات الجديدة القادمة. يتم تحديث الحالة المخفية باستخدام العلاقات التالية.

تعطى الحالة الحالية بالعلاقة (3-19) [40]:

$$h_t = f(h_{t-1} - x_t) \quad (19-3)$$

حيث أن: h_t : الحالة الحالية، h_{t-1} : الحالة السابقة، x_t : الدخل.

ومن ثم نطبق دالة التنشيط $\tanh$ كما في العلاقة (3-20)[40]:



الشكل (3-8) بنية كل من الشبكات العصبونية التكرارية (a) والشبكات العصبونية الأمامية (b) [40]

$$h_t = \tanh(w_{hh} h_{t-1} - w_{xh} x_t) \quad (20-3)$$

حيث أن w_{hh} : هو وزن العصبون الحالي، w_{xh} : هو وزن عصبون الدخل، يصبح الخرج كما في العلاقة (21-3) [40]:

$$y_t = w_{hy} h_t \quad (21-3)$$

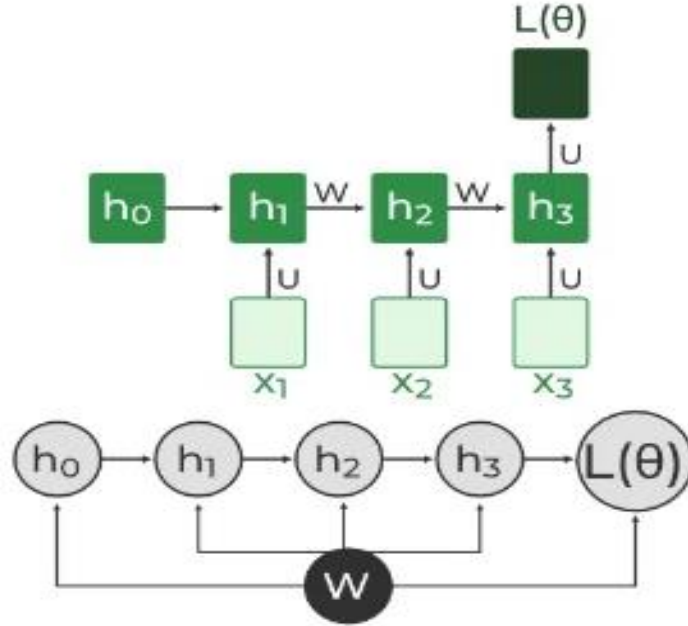
حيث w_{hy} : هو وزن عصبون الخرج.

إن خوارزمية الانتشار الخلفي المطبقة في RNN مرتبطة بالزمن وتسمى الانتشار العكسي عبر الزمن BPTT (backpropagation through time).

1.6.3 الانتشار الخلفي عبر الزمن BPTT

في RNN تكون الشبكة العصبونية منظمة، وبما أن الشبكة المنظمة تحسب كل متغير على حدة بترتيب محدد،

مثل h_1 أولاً ثم h_2 ثم h_3 وهكذا



الشكل (3-9) بنية الشبكة العصبونية التكرارية بعد فك الحلقات

سنطبق الانتشار العكسي على كل الحالات الزمنية المخفية هذه بشكل متسلسل:

تعتمد $L(\theta)$ (دالة الخسارة) على h_3

h_3 تعتمد بدورها على h_2 و W

h_2 تعتمد بدورها على h_1 و W

h_1 تعتمد بدورها على h_0 و W

حيث h_0 هي حالة بداية ثابتة، تعطى قيمة الخطأ بالعلاقة (22-3) [41]:

$$\frac{\partial L(\theta)}{\partial W} = \sum_{t=1}^T \frac{\partial L(\theta)}{\partial W} \quad (22-3)$$

لتبسيط المعادلة نقوم بتطبيق الانتشار الخلفي على سطر واحد فقط كمايلي: $\frac{\partial L(\theta)}{\partial W} = \frac{\partial L(\theta)}{\partial h_3} \frac{\partial h_3}{\partial W}$

إن الجزء $\frac{\partial L(\theta)}{\partial h_3}$ حسابه مثل حساب مشتق الخطأ للشبكات العادية

وبالتالي سوف نحسب الآن الانتشار الخلفي بالنسبة للشق الثاني $\frac{\partial h_3}{\partial W}$:

$$\frac{\partial h_3}{\partial W} = \frac{\partial h_3}{\partial W} + \frac{\partial h_3}{\partial h_2} \left[\frac{\partial h_2}{\partial W} + \frac{\partial h_2}{\partial h_1} \frac{\partial h_1}{\partial W} \right] = \frac{\partial h_3}{\partial W} + \frac{\partial h_3}{\partial h_2} \frac{\partial h_2}{\partial W} + \frac{\partial h_3}{\partial h_2} \frac{\partial h_2}{\partial h_1} \frac{\partial h_1}{\partial W} = \frac{\partial h_3}{\partial W} + \frac{\partial h_3}{\partial h_2} \frac{\partial h_2}{\partial h_1} \left[\frac{\partial h_1}{\partial W} \right] \frac{\partial h_2}{\partial W}$$

$$\frac{\partial h_3}{\partial h_2} \frac{\partial h_2}{\partial h_1} \left[\frac{\partial h_1}{\partial W} \right] \frac{\partial h_2}{\partial W} = \frac{\partial h_3}{\partial W} + \frac{\partial h_3}{\partial h_2} \frac{\partial h_2}{\partial W} + \frac{\partial h_3}{\partial h_1} \frac{\partial h_1}{\partial W}$$

وبالتالي تنتهي المعادلة إلى صيغة المجموع الذي يعطى بالعلاقة (23-3) [41]:

$$\frac{\partial h_3}{\partial W} = \sum_{k=1}^3 \frac{\partial h_3}{\partial h_k} \frac{\partial h_k}{\partial W} \quad (23 - 3)$$

تصبح علاقة تغير الخطأ (24-3) [41]:

$$\frac{\partial L(\theta)}{\partial W} = \frac{\partial L(\theta)}{\partial h_3} \sum_{k=1}^3 \frac{\partial h_3}{\partial h_k} \frac{\partial h_k}{\partial W} \quad (24-3)$$

2.6.3 المشكلات التي تواجه الشبكات العصبونية التكرارية

التدرج المتلاشي: تعد مشكلة التدرج المتلاشي واحدة من التحديات الرئيسية التي تجعل تدريب الشبكات العصبونية التكرارية (RNNs) أمراً صعباً. تحدث هذه المشكلة عندما يصبح المشتق صغيراً جداً أثناء عملية الانتشار العكسي (Backpropagation)، مما يؤدي إلى انعدام تأثيره الفعلي على تحديث الأوزان. هذا يعني أن الشبكة تواجه صعوبة في تعلم العلاقات طويلة الأمد في البيانات الزمنية أو التسلسلات [40].

التدرج المتفجر: على العكس من التدرج المتلاشي، يحدث التدرج المتفجر عندما تنمو التدرجات بشكل كبير أثناء التدريب. هذا النمو المفرط يؤدي إلى تحديثات ضخمة جداً للأوزان، مما يجعل الشبكة غير مستقرة وقد يؤدي إلى فشل التدريب. السبب الرئيسي لهذه المشكلة هو تراكم أخطاء كبيرة أثناء عملية الانتشار العكسي في الشبكات العصبونية.

أحد الحلول المقترحة للتغلب على مشاكل RNN البسيطة هو استخدام شبكة الذاكرة طويلة قصيرة المدى.

7.3 شبكة الذاكرة الطويلة قصيرة الأمد (LSTM) Long Short Term Memory

هي نسخة محسنة من الشبكات العصبية التكرارية (RNNs) التي تتمتع بقدرة على تعلم التبعية طويلة المدى في البيانات المتسلسلة. ذلك يعود إلى وجود خلية الذاكرة داخل الشبكة، والتي تعمل كحاوية يمكنها الاحتفاظ بالمعلومات لفترات زمنية طويلة، مما يساعد في التغلب على مشكلة التدرج المتلاشي.

تحتوي هندسة LSTM على بنية متسلسلة تحتوي على أربع شبكات عصبية وكتل ذاكرة مختلفة تسمى الخلايا. حيث يوضح الشكل (3-10) بنية LSTM يتم الاحتفاظ بالمعلومات بواسطة الخلايا ويتم إجراء عمليات معالجة الذاكرة بواسطة البوابات

• بوابة النسيان forget gate.

يتم إزالة المعلومات التي لم تعد مفيدة في الخلية باستخدام بوابة النسيان. يتم تغذية المدخلين X_t (المدخل في الزمن t) و h_{t-1} (مخرجات الخلية السابقة) إلى البوابة وضربهما بمصفوفات الوزن متبوعة بإضافة الانحياز. يتم تمرير الناتج من خلال دالة تنشيط تعطي خرجًا ثنائيًا.

إذا كان الخرج لحالة خلية معينة 0، يتم نسيان قطعة المعلومات وبالنسبة للخرج 1، يتم الاحتفاظ بالمعلومات للاستخدام في المستقبل.

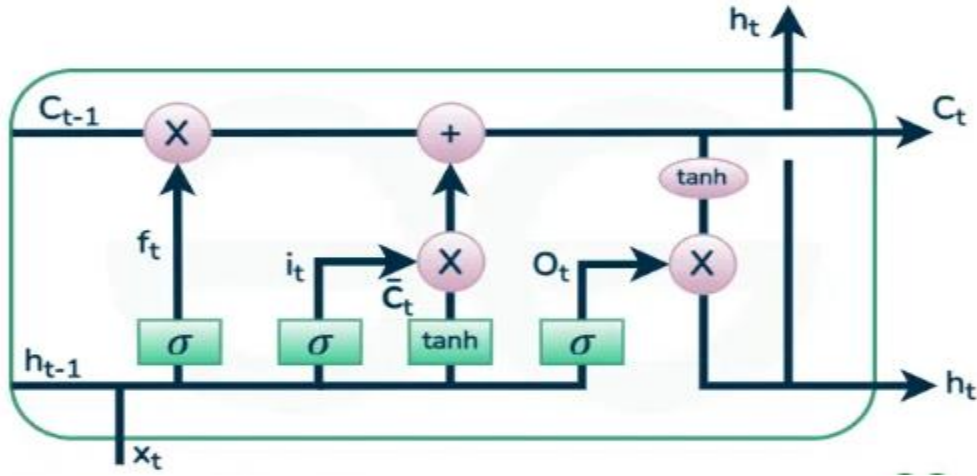
معادلة بوابة النسيان تعطى بالعلاقة (3-25):[42]:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (25-3)$$

حيث أن W_f تمثل مصفوفة الأوزان المرتبطة ببوابة النسيان.

$[h_{t-1}, x_t]$ تشير إلى تسلسل الإدخال الحالي والحالة المخفية السابقة.

b_f هي الانحياز الخاص ببوابة النسيان، σ هي دالة التنشيط.



الشكل (10-3) بنية LSTM [38]

• بوابة الإدخال

تتم إضافة المعلومات المفيدة إلى حالة الخلية بواسطة بوابة الإدخال. أولاً، يتم تنظيم المعلومات باستخدام دالة سيجمويد وتصفية القيم المراد تذكرها على غرار بوابة النسيان باستخدام المدخلات h_{t-1} و x_t . ثم يتم إنشاء متجه باستخدام دالة $\tanh$ التي تعطي خرجاً من -1 إلى $+1$ ، والذي يحتوي على جميع القيم الممكنة من h_{t-1} و x_t . وأخيراً، يتم ضرب قيم المتجه والقيم المنظمة للحصول على المعلومات المفيدة. معادلة بوابة الإدخال تعطى بالعلاقين (26-3) (27-3) [42]:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (26-3)$$

$$\hat{C}_t = \tanh(W_{\hat{C}} \cdot [h_{t-1}, x_t] + b_{\hat{C}}) \quad (27-3)$$

ومن ثم نوجد الحالة الجديدة C_t وتعطى بالعلاقة (28-3) [42]

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \hat{C}_t \quad (28-3)$$

• بوابة الإخراج

تتم مهمة استخراج المعلومات المفيدة من حالة الخلية الحالية بواسطة بوابة الإخراج، حيث يتم أولاً إنشاء متجه من خلال تطبيق دالة $\tanh$ على حالة الذاكرة. ثم تُنظم هذه المعلومات باستخدام دالة sigmoid التي تحدد القيم التي يجب تذكرها استناداً إلى المدخلات. في النهاية، يتم ضرب قيم المتجه الناتج عن دالة $\tanh$ مع القيم المنظمة بواسطة sigmoid ، ليتم إرسال النتيجة كمخرجات ومدخلات إلى الخلية التالية. معادلة بوابة الإخراج تعطى بالعلاقة (29-3)[42]:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (29-3)$$

8.3 المعلمات والمعلمات الفائقة

• **المعلمات (Parameters):** هي القيم الداخلية بالنسبة للنموذج والتي يتم تعلمها أو تقديرها من البيانات أثناء التدريب حيث تحاول الخوارزمية المستخدمة تعلم العلاقة بين ميزات الإدخال والأهداف، يبدأ تدريب النموذج عادةً بتهيئة المعلمات إلى بعض القيم (قيم عشوائية صغيرة). ومع تقدم التدريب، يتم تحديث القيم الأولية باستخدام خوارزمية التعلم التي تعمل على تحديث قيم المعلمات باستمرار مع تقدم التعلم [43]. مثل الأوزان (weights) والانحياز (bias).

• **المعلمات الفائقة (hyperparameters):** هي المعلمات التي تتحكم في عملية التعلم وتحدد قيم معلمات النموذج وهذه المعلمات تحدد قيمها أو تختار تكوينها قبل بدء التدريب وتبقى نفسها عند انتهاء التدريب.

بعض الأمثلة (نسبة تقسيم التدريب والاختبار، معدل التعلم (learning rate)، اختيار خوارزمية التعلم، اختيار دالة التنشيط، اختيار دالة التكلفة (loss function)، عدد الطبقات المخفية (hidden layer)، عدد وحدات التنشيط في كل طبقة، معدل التسرب (dropout probability)، عدد دورات التدريب (epochs) حجم النواة أو المرشح وحجم الحزمة (Batch).

9.3 استنظام الدفعة (Batch Normalization)

يعد استنظام الدفعة تقنية تستخدم في التعلم العميق لتسريع عملية التدريب وجعل الشبكات العصبونية أسرع وأكثر ثباتاً. تتم هذه العملية عن طريق ضبط تنشيط الطبقات في الشبكة العصبونية لكل حزمة صغيرة من البيانات أثناء التدريب. حيث تساعد هذه التقنية على تقليل قيمة تحول المتغير الداخلي [44]. تحول المتغير الداخلي (Internal Covariate Shift) هو التغير في توزيع تنشيط الشبكة بسبب التغير في معلمات الشبكة أثناء التدريب. أيضاً تساعد هذه التقنية على تحسين الاستقرار في الشبكة بالإضافة إلى تسريع تدريبها وجعل الشبكات العصبونية أسرع وأكثر ثباتاً. وذلك من خلال تسوية أو معالجة مدخلات الطبقات عن طريق إعادة التمرکز وإعادة التوسع.

تعتمد هذه التقنية على حساب متوسط مجموع قيم X للطبقة المطبق عليها μ بالعلاقة (30-3) [28]، ثم حساب التباين σ بالعلاقة (31-3) [28] ثم إجراء تنعيم ل X بالعلاقة (32-3) [44] والتي تعطي مخرجات نسميها $\hat{X}$ ثم بعد ذلك حساب قيمة ما يسمى Y وتساوي قيمة $\hat{X}$ مضروبة بغاما ومضاف لها قيمة بيتا. وهي التي تمثل القيمة النهائية) أي أن Y التي تعطى بالعلاقة (33-3) [44] هي القيمة التي سنستبدل بها قيمة X الأصلية. يتم حساب ناتج التقييس لبعض وحدات البيكسل المتمركزة في X على النحو التالي حيث m عدد العينات:

$$\mu = \frac{1}{m} \sum_{i=1}^m X_i \quad (30 - 3)$$

$$\sigma = \frac{1}{m} \sum_{i=1}^m (X_i - \mu)^2 \quad (31 - 3)$$

$$\hat{X} = \frac{X - \mu}{\sqrt{\sigma^2 - \mu}} \quad (32 - 3)$$

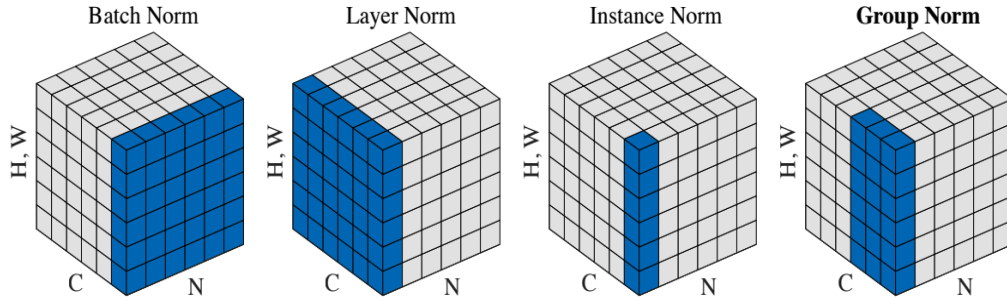
$$Y = \lambda * \hat{X} + \beta \quad (33-3)$$

λ, β بارامترات نختارها بقيم عشوائية لأن الشبكة العصبونية تعاملهما معاملتان الأوزان وبالتالي تقوم بتعديل

قيمتها من خلال خوارزمية الانتشار الخلفي

أنواع الاستنظام: يوجد لدينا أربع أنواع من التقييس حسب موقع إضافة التقييس كما هو موضح في الشكل (5-3)

(3) وهي: استنظام الدفعة، استنظام الطبقة، استنظام المثل، استنظام المجموعة.



الشكل (3-5) أنواع التقييس [30]

10.3 شبكة U-Net

U-Net عبارة عن بنية تعلم عميق مستخدمة على نطاق واسع كان الغرض الأساسي منها تجزئة الصور

تمكنت U-Net من معالجة تحدي البيانات المحدودة في المجال الطبي. تم تصميم هذه الشبكة للاستفادة بشكل

فعال من كمية أقل من البيانات مع الحفاظ على السرعة والدقة. سميت بهذا الاسم بسبب أن البنية تأخذ شكل حرف

U.

تتألف الشبكة الموضحة بالشكل (3-11) من المراحل التالية:

- المسار الانكماش أو مسار الترميز: في U-Net هذا المسار مسؤول عن تحديد الميزات المهمة في طبقة

الإدخال [45]. تقوم طبقات الترميز بإجراء عمليات التقافية تقلل من الدقة المكانية لخرائط الميزات ولكنها تزيد

من عمقها (يقل حجم الصور المدخلة ولكن يزداد عددها)، وبالتالي النقاط تمثيلات تجريدية بشكل متزايد

للمدخلات. يشبه مسار الانكماش هذا طبقات التغذية الأمامية في الشبكات العصبونية التلافية الأخرى.

- **المسار التمديدي أو مسار فك التشفير:** هذا المسار مسؤول على فك تشفير وتحديد موقع الميزات مع الحفاظ على الدقة المكانية للمدخلات. تعمل طبقات فك التشفير في المسار التمديدي على زيادة حجم خرائط الميزات، عن طريق ما يسمى عملية المنقول *transpose*.
- **اتصال التخطي (skip connection):** يعيد تقديم الميزات التفصيلية إلى مسار فك التشفير والتي كانت قد فقدت أثناء مرحلة التشفير وذلك عبر عملية *concatenating* التي تربط الميزات مع بعضها البعض.

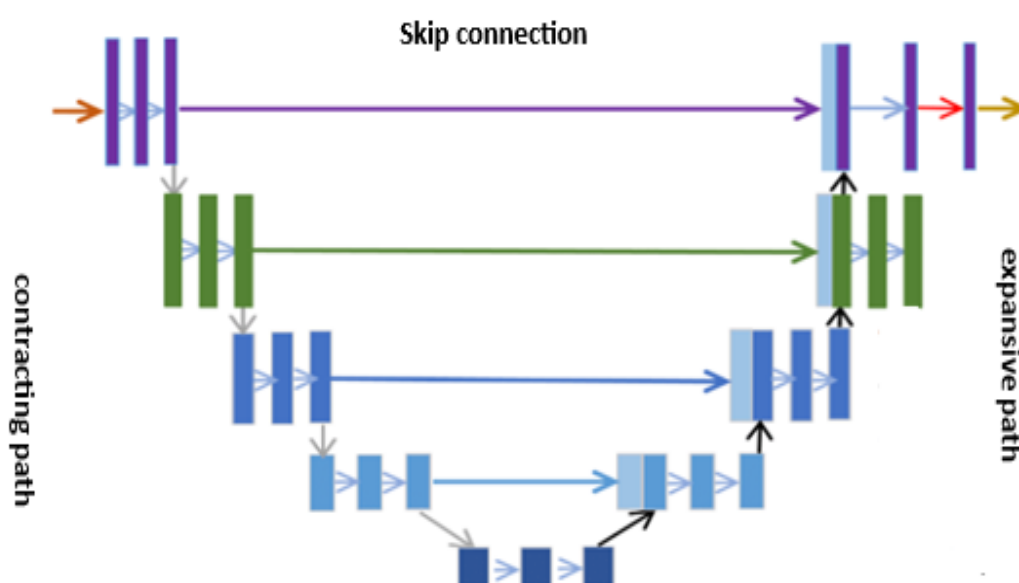
11.3 تكثير الصورة

تكثير الصورة (Image Augmentation) IA هو تقنية لزيادة حجم مجموعة التدريب بشكل مصطنع عن طريق إنشاء نسخ معدلة من الصور الموجودة باستخدام البيانات الموجودة في مجموعة التدريب. تُعد هذه التقنية شائعة كإجراء لتخفيف الإفراط في التدريب في نماذج التعلم العميق وتُستخدم على نطاق واسع لتحسين الأداء. من خلال تكثير الصورة، يتم توليد أمثلة تدريبية جديدة تشبه الأصلية عبر سلسلة من التغييرات العشوائية على الصور، مما يساعد على توسيع حجم مجموعة التدريب [46].

تشمل التغييرات التي يتم إجراؤها على الصور التحويلات الهندسية و التحويلات اللونية مثل الانقلاب، تغيير الحجم، القص، تعديل السطوع، التباين، أو حتى حذف جزء من الصورة الأصلية أو دمج عدة صور معًا. تُحفز هذه التقنية بفكرة أن التعديلات العشوائية على أمثلة التدريب تتيح للنموذج الاعتماد بشكل أقل على سمات محددة، وبالتالي تحسين قدرته على التعميم. على سبيل المثال، قد يتم اقتصاص الصورة بطرق مختلفة لعرض الكائن في مواقع متنوعة، مما يقلل من اعتماد النموذج على موضع الكائن. كما يمكن ضبط عوامل مثل السطوع واللون لتقليل حساسية النموذج للألوان. يستخدم تكثير الصور لمنع النماذج من الإفراط في الملاءمة وفي حالة كانت مجموعة التدريب الأولية صغيرة جداً وبالتالي تحسين دقة النموذج كما أنها تستخدم لتقليل التكلفة التشغيلية لعنونة البيانات وتنظيف مجموعة البيانات الخام.

12.3 نقل التعلم Transfer Learning

هو تقنية تعلم آلي يتم فيها استخدام المعرفة المكتسبة من خلال مهمة أو مجموعة بيانات واحدة لتحسين أداء النموذج في مهمة أخرى ذات صلة أو مجموعة بيانات مختلفة. يستخدم نقل التعلم ما تم تعلمه في مهمة محددة لتحسين التعميم في مهمة أخرى [47]، يوجد العديد من التطبيقات التي تستخدم نقل التعلم، من حل مشاكل الانحدار في علم البيانات إلى تدريب نماذج التعلم العميق. والواقع أنه جذاب بشكل خاص للتعلم العميق نظراً للكمية الكبيرة من البيانات اللازمة لإنشاء شبكات عصبية عميقة.



الشكل (3-11) بنية U-Net [40]

تقوم عمليات التعلم التقليدية ببناء نموذج جديدًا لكل مهمة جديدة، استنادًا إلى مجموعة البيانات المتاحة، وذلك لأن خوارزميات التعلم الآلي التقليدية تفترض أن بيانات التدريب والاختبار تأتي من نفس مساحة الميزة، وبالتالي إذا تغير توزيع البيانات، أو تم تطبيق النموذج المدرب على مجموعة بيانات جديدة، فيجب على المستخدمين إعادة تدريب نموذج جديد من الصفر، حتى لو كانت المهمة مماثلة للمهمة في النموذج الأول بالمقابل، فإن خوارزميات نقل التعلم تأخذ النماذج أو الشبكات المدربة مسبقاً والتي تم تخزين أوزانها كنقطة بداية للتدريب ، ثم

تطبق المعرفة التي اكتسبها هذا النموذج في مهمة أو بيانات سابقة نحو مهمة أو بيانات مستهدفة جديدة ذات صلة بالمهمة القديمة.

13.3 البيانات المستخدمة في التعرف على المشي

تتنوع بيانات الإدخال المستخدمة في أنظمة التعرف على المشي لتشمل أنواعًا متعددة مثل الفيديو، وصور RGB ، والصور الظلية، وصور طاقة المشية (GEI) ، وصور التدفق البصري، وهيكل الجسم (Skeleton) ، والبيانات المستخلصة من أجهزة الاستشعار المختلفة. يمكن لأجهزة الاستشعار القابلة للارتداء أو تلك المدمجة في الهواتف الذكية جمع بيانات متنوعة تشمل الحركة، والضغط، والسرعة، والتسارع.

في هذا البحث، سيتم استخدام بيانات الصور بالإضافة إلى بيانات الحساسات المدمجة في الهواتف الذكية والمثبتة على الحزام، بالاعتماد على مجموعات بيانات جُمعت مسبقًا لتحقيق هدف التعرف على المشي.

1.13.3 البيانات المرئية (الصور والفيديو)

غالبًا ما يعتمد تحليل المشية على التقدم الزمني لحركات الشخص، وحتى عندما يتم التقاط الصور في أوقات مختلفة، فإنها تمثل مراحل مختلفة من دورة المشية. وبالتالي، فإن مجموعة بيانات الصور المأخوذة من العينات لا تزال تغلف سمات المشية الأساسية اللازمة للتحديد الدقيق. علاوة على ذلك، يوفر استخدام الصور من مقاطع الفيديو الاختبار من جهات متعددة وظروف بيئية متنوعة، مما يعزز من قوة النموذج، تحافظ هذه الطريقة على سلامة سمات المشية دون الحاجة إلى تحليل فيديو مستمر، مما يحسن الموارد الحسابية مع ضمان تحديد موثوق للشخص. مجموعة بيانات Market مكونة من صور مستخرجة من تسجيلات فيديو في أوقات مختلفة للبيئة، هذا النهج صالح علميًا ومنطقيًا لأن كل صورة تم التقاطها في نقاط زمنية مميزة تحافظ على معلومات مهمة تتعلق بالمشية.

1.1.13.3 مجموعة البيانات Market (Market dataset)

تحتوي مجموعة البيانات Market على صور تم التقاطها أمام سوبر ماركت تابعة لجامعة تسينغهاو (Tsinghua University) حيث تم استخدام ست كاميرات (5 كاميرات بدقة عالية وواحدة بدقة منخفضة) في بيئة عشوائية تحتوي على 1501 هوية وعلى أكثر من 12900 صورة حيث يتواجد كل شخص في صورتين لكاميرتين على الأقل [48] ويحتوي اسم الصورة على: رقم الشخص ورقم الكاميرا ورقم تسلسل الصور داخل المشهد وهو رقم عشوائي.

مثال : إذا كان اسم الصورة 0001c1s100230100 فإن رقم الشخص هو 0001 والكاميرا c1 وصورة المشهد s1.

2.13.3 البيانات الناتجة عن الحساسات

لقد مهد التطور السريع للأنظمة الكهروميكانيكية الدقيقة Micro Electro Mechanical Systems (MEMS) الطريق لبعض الاختراعات العلمية والتطبيقية الهامة في العديد من مجالات البحث [49]، حيث تعتبر أجهزة الاستشعار بالقصور الذاتي (مقاييس التسارع، الجيروسكوبات) التي يتم دمجها معًا بشكل شائع كوحدات القياس العطالية (IMU) أحد أهم أعضاء عائلة (MEMS) نظرًا للعديد من الخصائص الإيجابية، مثل الوزن الخفيف، والحجم الصغير، وانخفاض استهلاك الطاقة، والقدرة على النقل والتكلفة المنخفضة، أصبحت أجهزة الاستشعار بالقصور الذاتي مستخدمة على نطاق واسع ولا غنى عنها في جميع الأنشطة التي تعالج الحركة بشكل غير مباشر أو مباشر. علاوة على ذلك، من المعروف منذ فترة طويلة أن البيانات التي يتم الحصول عليها بواسطة أجهزة الاستشعار بالقصور الذاتي يمكن معالجتها من خلال مناهج متقدمة من أجل إجراء تحليل حركة معقد. يتم تكريس اهتمام خاص لمشكلة تحليل المشي لأن القدرة على المشي تمثل واحدة من أهم الوظائف الحيوية التي لها تأثير كبير على نوعية الحياة.

في الوقت الحاضر، أصبح إجراء تحليل المشي باستخدام أجهزة استشعار بالقصور الذاتي أمرًا لا غنى عنه في العديد من مجالات البحث بما في ذلك الميكانيكا الحيوية وإعادة التأهيل العصبي والطب الرياضي [50]

، يمكن لأجهزة الاستشعار بالقصور الذاتي قياس مسارات الحركة أحادية أو متعددة النقاط لأجزاء جسم واحدة أو متعددة من الشخص أثناء المشي. أثناء فترة القياس، يتم الحصول على إشارات توفر معلومات فورية عن الكمية المقاسة (أي التسارع المكاني عند استخدام مقياس التسارع ثلاثي المحاور). وبهذه الطريقة، يمكن تقييم مشية الشخص من حيث معلومات المشي التي يمكن تفسيرها بعدة طرق من أجل اكتشاف أو ملاحظة ظاهرة معينة من ناحية أخرى، ونظرًا لحقيقة أن كل فرد يمتلك طريقة فريدة للمشي، يمكن استغلال تقييم المشية بالاعتماد على أجهزة استشعار بالقصور الذاتي بطريقة مماثلة لمشكلة التعرف على المشية عن طريق البيانات المرئية. وبهذه الطريقة، يمكن تفسير المشية كسمة حيوية، وبالتالي، تتمتع أجهزة الاستشعار بالقصور الذاتي بإمكانات كبيرة للعب دور مهم في مجال القياسات الحيوية. وبالتالي، فإن تطبيق مثل هذا النهج الحيوي يمكن أن يعزز بشكل كبير جوانب الأمن التي يمكن تمثيلها من خلال العديد من سيناريوهات الاستخدام، بما في ذلك إجراء تحقق جديد يمكنه توسيع أو حتى استبدال آليات الأمان الحالية واكتشاف السرقة، وتبديل الملف الشخصي، وتتبع المستخدم، ودعم أنظمة الرعاية الصحية المتنقلة، وغيرها الكثير.

إن تطوير أساليب التعرف على المشية باستخدام أجهزة الاستشعار بالقصور الذاتي قد تزامن مع الانتشار الواسع للأجهزة الذكية مثل الساعات الذكية [51] ، والنظارات، والهواتف الذكية، والأجهزة اللوحية، وغيرها. أدى ذلك إلى شيوع دمج أجهزة الاستشعار بالقصور الذاتي في هذه الأجهزة الذكية. هناك حقيقتان بارزتان تسلطان الضوء على أهمية أجهزة الاستشعار بالقصور الذاتي في تحليل المشي:

أولاً، هناك قاعدة كبيرة من المستخدمين الذين يمتلكون ويستخدمون الأجهزة الذكية بشكل يومي، مما يوفر مصدرًا غنيًا ومتاحًا للبيانات. ثانيًا، تتميز أجهزة الاستشعار بالقصور الذاتي المدمجة في الأجهزة الذكية بقدرتها على التغلب على العديد من القيود التقليدية مثل كفاءة الطاقة، والقوة الحاسوبية، وعرض النطاق الترددي، والتكلفة، وسعة التخزين.

أثبتت الدراسات أن البيانات المستخلصة من هذه الأجهزة يمكن استخدامها لتقييم حركة المستخدمين بدقة، بما في ذلك التعرف على النشاط وتحليل الحركة، مما يجعل استخدام الأجهزة الذكية المعتمدة على بيانات القصور الذاتي أداة معقولة وفعالة في التعرف على المشي. وقد أدى ذلك إلى تصميم العديد من الأدوات لقياس هذه البيانات، مثل وحدة القياس العطالية (IMU)، التي أصبحت عنصراً مهماً في هذا المجال.

1.2.13.3 وحدة القياس العطالية (Inertial Measurement Unit)

أجهزة قياس التسارع والجيروسكوبات IMU هي أجهزة استشعار بالقصور الذاتي نموذجية ومضمنة بمعظم الهواتف الذكية، تقيس أجهزة قياس التسارع والجيروسكوبات ديناميكيات القصور الذاتي في ثلاث اتجاهات، وهي على طول المحاور X و Y و Z [52]. يعتمد مقياس التسارع ثلاثي المحاور على المبدأ الأساسي للتسارع ويُستخدم لقياس تسارع الهاتف الذكي (بما في ذلك الجاذبية) في اتجاهات X و Y و Z. تعكس التسارعات في الاتجاهات الثلاثة التغيرات في السرعة الخطية للهاتف الذكي في الفضاء ثلاثي الأبعاد، وبالتالي تعكس حركة مستخدم الهاتف الذكية. يلتقط الجيروسكوب ثلاثي المحاور السرعة الزاوية للهاتف الذكي أثناء دورانه في الفضاء، ويمكن استخدامه أيضاً لوصف نمط حركة المستخدم.

في هذا البحث، يتم استخدام هواتف ذكية من إنتاج Samsung و Xiaomi و Huawei، جميعها تعمل بنظام تشغيل Android. أثناء المشي، يتسارع الهاتف الذكي ويدور متأثراً بحركات المستخدم. تُعد هذه البيانات فريدة لكل فرد، ما يجعلها مصدراً موثقاً لتحليل ديناميكيات المشي والتعرف على أنماطه.

يمكن أن تؤدي بيانات المشية المقسمة حسب الدورة غالباً إلى تحسين الأداء مقارنة بالبيانات غير المقسمة حسب الدورة، لذلك من الضروري تقسيم بيانات المشية المستخرجة حسب خطوات المشي. بمجرد استخراج بيانات المشية تقسم البيانات المستمرة إلى خطوات منفصلة. والجدير بالذكر أن الخطوة الكاملة تشير إلى لمس الأرض بنفس القدم مرتين على التوالي [53].

الهدف الأساسي لهذا النظام هو الاستفادة من أجهزة استشعار الهواتف الذكية لتحسين قدرات استخراج

المشية والتعرف عليها، من خلال الاستفادة من أجهزة الاستشعار المضمنة في الهواتف الذكية، مثل مقاييس التسارع وأجهزة قياس الزوايا، يهدف النظام إلى تقديم حل ملائم وغير تدخلية لتحليل المشية، والذي يحمل إمكانات كبيرة في تطبيقات مختلفة بما في ذلك الرعاية الصحية، والقياسات الحيوية، والتفاعل بين الإنسان والحاسوب.

الجيروسكوب هو نوع من أجهزة الاستشعار، والذي يشير بدلاً من الإشارة إلى الاتجاه، إلى معدل تغير زاوية الجسم مع الوقت. تقيس الجيروسكوبات متجه السرعة الزاوية لجهاز الاستشعار أو للجسم المثبت عليه [54]. مقاييس التسارع هي أجهزة استشعار تقيس التسارع المطبق على نقطة معينة من الجسم مثل مركز كتلة الجسم المادي حيث يُفترض أن توجد وحدة القياس العطالية [54].

تقدير زوايا الجسم [52] هو مسألة مهمة ويسمى تقدير الموضع؛ يشير الموضع إلى الزوايا الثلاث الرئيسية لأي جسم متحرك. الزوايا الثلاث هي التدرج والانحدار والانحراف. يجب تثبيت المستشعرات على مركز كتلة الجسم المستهدف، لنفترض أن الجسم بشري وأن المستشعرات تقع على الرأس، فإن الزوايا الثلاث ستكون كنقطة أخرى، وستكون الزوايا مختلفة. تمثل مخرجات IMU بصمة مشية الشخص والتي يتم تلخيصها على أنها تفاصيل الحركة والميزات التي تعكس مشية كل شخص، يتم استخراج بيانات الحساسات ومعالجتها وتنظيمها ضمن مجموعات بيانات مجدولة مثل مجموعتا البيانات OU-ISIR و whuGAIT التي سنستخدمها في هذا البحث.

2.2.13.3 مجموعة بيانات whuGAIT: تحتوي مجموعة بيانات whuGAIT على بيانات عطالية لـ 118 شخصاً تم جمعها بواسطة الهواتف الذكية [52] وتتكون من 8 مجموعات بيانات فرعية. مجموعات البيانات من 1 إلى 4 مخصصة لتحديد هوية الشخص، ومجموعات البيانات 5 و 6 مخصصة للمصادقة. تُستخدم مجموعات البيانات 7 و 8 لفصل بيانات المشي عن بيانات عدم المشي. من بين 118 شخصاً، يجمع 20 شخصاً البيانات في يومين، مع آلاف العينات لكل منهم، ويجمع 98 شخصاً آخرين البيانات في يوم واحد، مع مئات العينات لكل منهم. تحتوي عينة البيانات على بيانات مقياس التسارع ثلاثي المحاور وبيانات الجيروسكوب ثلاثي المحاور. معدل أخذ العينات للمستشعرات هو 50 هرتز.

3.2.13.3 وصف مجموعة البيانات OU-ISIR: مجموعة بيانات OU-ISIR هي قاعدة بيانات متخصصة

تحتوي على بيانات من أجهزة الاستشعار بالقصور الذاتي، وهدفها الأساسي دعم الجهود البحثية المتعلقة بتطوير واختبار خوارزميات تحديد هوية الأفراد استنادًا إلى أنماط المشي. يتم توزيع هذه القاعدة بواسطة معهد البحث

العلمي والصناعي (OU-ISIR) بجامعة أوساكا [55]، الذي يحتفظ بحقوق الطبع والنشر الخاصة بها. تم جمع البيانات خلال برنامج توعوي استمر لمدة خمسة أيام في طوكيو، اليابان، عام 2011. وافق جميع المشاركين في الدراسة على تقديم بياناتهم من خلال توقيع نموذج موافقة رسمي، بالإضافة إلى تزويدهم بمعلومات بيولوجية أساسية مثل العمر والجنس. تُعتبر هذه القاعدة أكبر مجموعة بيانات في العالم متخصصة في تحليل المشي باستخدام أجهزة الاستشعار بالقصور الذاتي، حيث تحتوي على بيانات لحوالي 741 فردًا أثناء حركتهم. تم استخدام أربعة أجهزة استشعار مثبتة على الحزام لجمع البيانات، مما يوفر مجموعة شاملة وقابلة للاستخدام في الأبحاث المتعلقة بالتعرف على المشي.

البيانات التي تم جمعها لم تكن كلها بنفس المستوى من حيث إمكانية استخدامها للتمييز بين الأفراد بعد القيام بعملية معالجة أولية أساسية لاستخراج البيانات المفيدة والتخلص من البيانات غير الصالحة. نتيجة لذلك، تم العامل مع البيانات على قسمين، القسم الأول وهو الأكثر جدوى ويحوي على بيانات مفيدة وواضحة وتم إنشاء المجموعة الأولية للبيانات بأكبر عدد ممكن من الأفراد. وفي الوقت نفسه، تم التعامل بطريقة أخرى مع إعطاء أهمية لتحسين هذه المجموعة مع تعديل نوع الحساس وموقعه الدقيق، وظروف الأرض، والموقع. سجل مستشعر IMUZ المركزي بيانات المشي على سطح مستوٍ لـ 744 شخصاً (389 رجلاً و355 امرأة) تتراوح أعمارهم بين 2 إلى 78 عاماً. باستخدام قيود مسار الحركة والارتباط الذاتي للإشارات، تم استخراج تسلسلين مميزين للمشبي على سطح مستوٍ لكل فرد بشكل تلقائي ضمن هذه المجموعة الفرعية.

ركزت المجموعة الفرعية الثانية بشكل أساسي على اختلافات موضع المستشعر وانحدار الأرض. استعدنا طريقتين للمشبي على سطح مستوٍ وعلى منحدر صاعد، وسلسلة أخرى للمشبي على منحدر هابط لكل فرد ولكل

مستشعر. تضمنت هذه المجموعة الفرعية 495 شخصاً تم جمع بياناتهم باستخدام ثلاث مستشعرات IMUZ ، و408 أشخاص تم جمع بياناتهم باستخدام الهاتف الذكي. في هذه المجموعة الفرعية، تم استخراج البيانات يدوياً عن طريق المزامنة مع مقطع فيديو تم تسجيله في الوقت نفسه. تم استخراج البيانات تلقائياً، ولكل هوية فردية هناك ملفان. تحتوي الصفوف الأولى من كل ملف على عدد الصفوف الموجودة في الملف؛ وهذا يشير إلى عدد العينات التي تم تسجيلها. على سبيل المثال، إذا كان هذا الرقم هو 600، فإن ذلك يعني أن الفرد قد مشى لفترة زمنية تعادل 600 عينة من القياسات.

الفصل الرابع

تنفيذ خوارزميات التعلم العميق للتعرف على الأشخاص باستخدام

بيانات المشي المرئية

Implementing Deep Learning Algorithms to

Recognize People Using Visual Gait data

الفصل الرابع

تنفيذ خوارزميات التعلم العميق للتعرف على الأشخاص باستخدام بيانات المشي

المرئية

1.4 مقدمة

بعد أن درسنا في الفصل الثالث الأسس النظرية لبعض خوارزميات التعلم العميق وتقنياته، سوف نخبر خوارزميتي CNN مع تقنية تكثير البيانات و ViT مع تقنية نقل التعلم على نفس مجموعة البيانات Market ونقارن النتائج التي حصلنا عليها.

2.4 البيئة البرمجية

تم استخدام Anaconda كبيئة برمجة. وفقاً للعديد من المراجع، فإن Anaconda هي بيئة مجانية ومفتوحة المصدر للغات البرمجة Python و R للحوسبة العلمية والتي تهدف إلى تبسيط إدارة الحزم ونشرها. يتضمن التوزيع حزم علم البيانات المناسبة لنظامي التشغيل Windows و Linux و macOS. تم تطويره وصيانته بواسطة Anaconda, Inc، التي أسسها بيتر وانج و ترافيس أوليفانت في عام 2012.

3.4 اختبار الشبكة العصبونية العميقة للحصول على أفضل بارامترات وتطبيق تقنية تكثير

البيانات

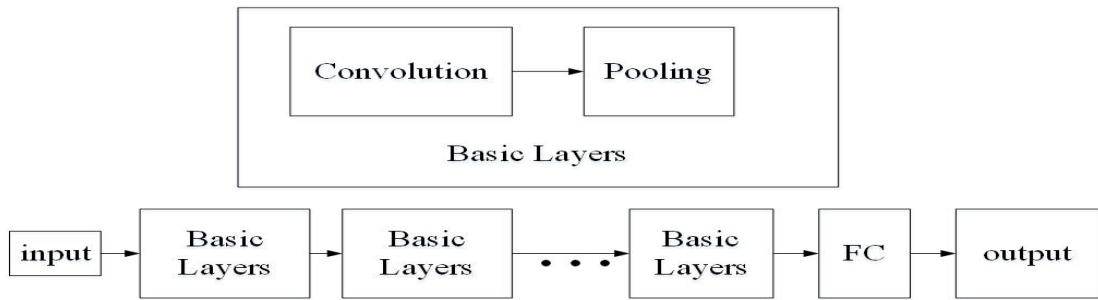
1.3.4 طبقات الشبكة العصبونية المقترحة

لهذه الدراسة تتكون CNN العميقة النموذجية من طبقة الإدخال، طبقات التلاف، طبقات الانتقاء، الطبقة المتصلة بالكامل (FC)، طبقة الإخراج. حيث أن طبقة الإدخال تتكون من البيانات الخام للصور الموجودة في مجموعة البيانات واللازمة لتغذية الشبكة العميقة.

في النموذج المقترح نقوم بتطبيق طبقة التلاف تليها طبقة التجميع (الطبقات الأساسية) الموضحة بالشكل

(1-4) ونقوم بتكرار هذه الطبقات بهدف الحصول على أفضل الميزات ومن ثم نطبق بقية الطبقات في نهاية النموذج. نعرف المصطلحات التالية: CL: طبقة التلاف، PL: طبقة التجميع، FC: الطبقة المتصلة بالكامل، يتكون النموذج المقترح من تعاقب الطبقات التالية:

Input → CL1 → PL1 → CL2 → PL2 → CL3 → PL3 → CL4 → PL4 → CL5
→ PL5 → FC1 → Dropout → FC2 → Regression.



الشكل (1-4) الطبقات الأساسية وتكرارها ضمن نموذج CNN المقترح (الباحثة)

نقوم بتهيئة الطبقات بقيم شائعة في المراجع والدراسات العلمية، تم تهيئة طبقة التلاف بالقيم الموجودة في الجدول (1-4) وتهيئة طبقة التجميع بالقيم الموجودة في الجدول (2-4) وتهيئة الطبقة المتصلة بالكامل بالقيم الموجودة في الجدول (3-4) وتهيئة طبقة الخرج كما في الجدول (4-4).

الجدول (1-4) تهيئة طبقة الالتفاف

CL_number	Filter_size	Kernel_size	Activation_funcon	Bias
CL ₁	32	3*3	ReLU	True
CL ₂	64	3*3	ReLU	True
CL ₃	128	3*3	ReLU	True
CL ₄	64	3*3	ReLU	True
CL ₅	32	3*3	ReLU	True

الجدول (4-2) تهيئة طبقات الانتقاء

Pl_number	Kernel_size	Pooling_Type
PL ₁₋₂₋₃₋₄₋₅	5*5	Max_Pooling

الجدول (4-3) تهيئة الطبقات التامة الربط

FC_number	Number of neurons	Activation_functon
FC ₁	1024	ReLU
FC ₂	200	ReLU

الجدول (4-4) تهيئة طبقة الإخراج

Regression	optimizer = adam	loss=categorical crossentropy
------------	------------------	-------------------------------

2.3.4 تحسين النموذج

تم استخدام 200 شخص من مجموعة البيانات Market، 3104 صورة لمجموعة التدريب و 840 صورة لمجموعة الاختبار، تم تشغيل 50 دورة فقط لأن عملية التدريب قد تستغرق وقتاً طويلاً، تم تطبيق تحسين النموذج عن طريق اختبار العديد من المصطلحات لاختيار أفضل قيم للمحاكاة.

إن عملنا الرئيسي يشبه إلى حد كبير عمل الخوارزمية الجينية. لكل مصطلح، قمنا باختبار بعض القيم ومراقبة الدقة والحقبة التي وصل فيها النموذج إلى دقة عالية. بالإضافة إلى ذلك، تم تنفيذ تكثير الصور Image Augmentation لتحسين دقة التعرف. تم إجراء تعديلات على التصميم والبنية لتحقيق أفضل دقة. تمت برمجة النموذج باستخدام مكتبة Tflern.

1.2.3.4 اختبار معدل التعلم: من أجل القيم السابقة للنموذج نقوم بتدريب النموذج وفي كل مرة نقوم بتبديل

قيمة معدل التعلم وإعادة تدريب النموذج حيث يرشدنا الى اختيار قيم نسبة التعلم الدراسات ونماذج CNN الشائعة. كما في الجدول (4-5).

الجدول (4-5) اختبار سرعة التعلم

<u>LR</u>	Accuracy
1e-5=0.00001	too bad
1e-4=0.00010	37%
5e-4=0.00050	55%
7e-4=0.00070	58.45%
8e-4=0.00080	59.17%
9e-4=0.00090	59%
1e-3=0.00100	56%
2e-3=0.00200	44%
5e-2=0.05000	Bad%

من هذه النتائج، أفضل قيمة لنسبة التعلم هي $LR = 8e-4 = 0.00080$ حيث LR هو رمز نسبة التعلم.

2.2.3.4 اختبار معامل الإسقاط (keep_prop): إن أحد الطبقات التي يمكن إضافتها لتحسين الأداء وتقليل

الإفراط في التدريب هي طبقة Dropout حيث يتم من خلالها إيقاف أو "إسقاط" بعض الوحدات العصبونية في مرحلة التدريب من خلال تحديد قيمة معامل

keep_prop الذي يحدد عدد الوحدات التي تبقى مفعلة، مما يجبر النموذج على تعلم تمثيلات متنوعة للبيانات.

نقوم الان باختيار قيمة نسبة التعلم المثلى $LR=8e-4=0.00080$ ونغير بقيمة معامل الاسقاط حتى نحصل على أفضل قيمة يوضح الجدول (4-6) القيم ونتائج الدقة التي تقابلها.

الجدول (4-6) اختيار أفضل قيمة لمعامل الإسقاط

Keep_Probability	Accuracy
0.3	59.29%
0.5	59.17%
0.7	63.57%
0.8	64.29%
1	55.12%

من هذه النتائج، أفضل قيمة لمعامل التسرب هي $Keep_Probability = 0.8$.

3.2.3.4 اختبار أفضل قيمة لأبعاد نواة طبقة التلاف: باختيار $LR = 8e-4$ ، $Keep_Probability =$

0.8 يوضح الجدول (4-7) قيم الدقة المقابلة لنواتي التلاف .

الجدول (4-7) قيم الدقة مقابل نواتي الالتفاف

Convolution_kernel_size	Accuracy
3*3	64.29%
5*5	63%

حجم نواة طبقة التلاف: القيمة 3 كانت الأفضل لحجم نواة التلاف Kernel_size_c

4.2.3.4 اختبار أفضل قيمة لأبعاد نواة الانتقاء: اختيار $LR = 8e-4$ ، $Keep_Probability = 0.8$ ، عن

طريق تغيير حجم النواة لطبقة الانتقاء $convolution_kernel_size=3*3$ ، ومراقبة دقة التحقق يوضح الجدول

(4-8) قيم الدقة المقابلة لنواتي الانتقاء .

الجدول (4-8) قيم الدقة مقابل نواتي الانتقاء

Pooling_kernel_size	Accuracy
3*3	63.69%
5*5	64.29%

من هذه النتائج أفضل قيمة لحجم نواة التجميع هي 5*5، لقد اختبرنا نوعي التجميع ولكن النوع المتوسط كان أفضل من الأعظمي

دالة الخسارة لطبقة الخرج: نختار (categorical cross-entropy) نظرًا لأن قوائم العناوين تمثل هويات الأشخاص والذي يبلغ عددهم 200 (ليس توقع ثنائي).

نضيف طبقات التقويس فيصبح هيكل النموذج كما هو موضح بالشكل (4-1) حيث نرمز لطبقة التقويس ب BN:

Input → CL1 → PL1 → BN1 → CL2 → PL2 → BN1 → CL3 → PL3 → BN1 → CL4 → PL4 → BN1 →

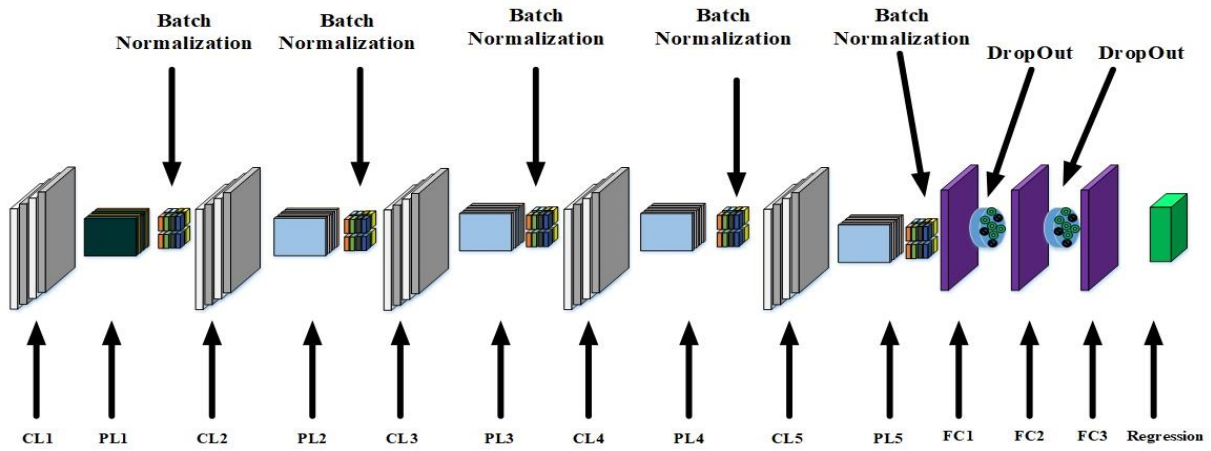
CL5 → PL5 → BN1 → FC1 → Dropout → FC2 → Dropout → FC3 → Regression

بعد تطبيق استنظام الدفعة تصبح الدقة 83%.

5.2.3 تكثير الصورة: استخدمنا دالة تكثير الصور التي تتضمن الميزات التالية:

- نطاق الدوران هو قيمة بالدرجات (0-180)، وهو نطاق يمكن من خلاله تدوير الصور عشوائيًا.
- نطاق القص لتطبيق تحويلات القص عشوائيًا.
- نطاق التكبير/التصغير للتكبير العشوائي داخل الصور.
- الانعكاس الأفقي هو لانعكاس نصف الصور أفقيًا عشوائيًا.
- نطاق السطوع: مجموعة أو قائمة من رقمين عائمين. نطاق لاختيار قيمة تحول السطوع.

إن تطبيق تكثير الصور على الصورة في الشكل (4-2) سينتج عنه ست صور كما في الشكل (4-3):



الشكل (1-4) بنية نموذج CNN المقترح لاختبار بيانات Market (الباحثة)



الشكل (2-4) صورة من مجموعة البيانات Market التي سيطبق عليها تكثير الصور

يكتب تابع التكثير في بايثون:

```
datagen = ImageDataGenerator (Rotation-range = 40,
```

```
Shear-range = 0.2,
```

```
Zoom-range = 0.2,
```

```
Horizontal-flip = True,
```

```
Brightness-range = (0.5, 1.5))
```



الشكل (4-3) ناتج عملية تكثير الصور

تصبح الدقة الآن بعد استخدام تكثير الصورة 96.23 %

4.4 تنفيذ خوارزمية محول الرؤية باستخدام تقنية نقل التعلم

سنستخدم تقنية نقل التعلم حيث سنأخذ نموذج vit-base-patch-16-224 المدرب مسبقاً كنقطة بداية ونقوم بتدريبه مرة أخرى على مجموعة البيانات Market.

1.4.4 نموذج vit-base-patch-16-224

هو نموذج محول مرئي مدرب مسبقاً على مجموعة البيانات ImageNet-21k، وهي مجموعة بيانات كبيرة تضم 14 مليون صورة و 21843 فئة، يأخذ الصور بحجم 224*224 كمدخلات [56] ، والتي يتم تقسيمها إلى رقع ذات حجم 16*16 ثابت كما في الشكل (4-4) وبالتالي فإن عدد الرقع يحسب كمايلي :

$$N = \frac{224 \times 224}{16 \times 16} = 196 \text{ patch}$$

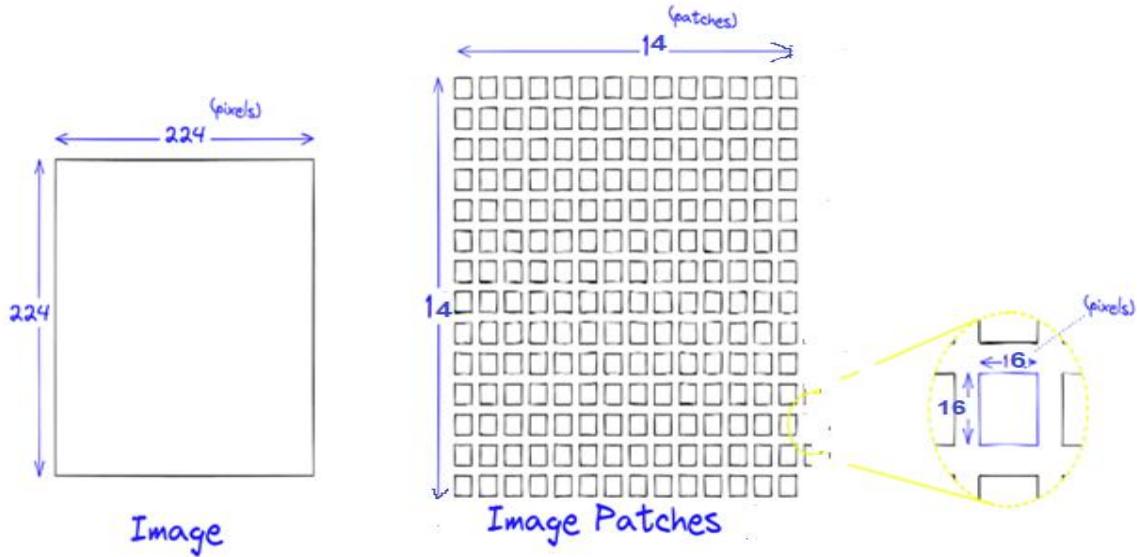
لتمثيل كل صورة ببعدين نأخذ الجذر التربيعي: $\sqrt{N} = \sqrt{196} = 14$

يتم تحويل الصورة إلى رقع باستخدام طبقة Conv2D. وكما نعلم، تقوم طبقة Conv2D بعمليات تلافية

ثنائية الأبعاد على بيانات الإدخال لتعلم الميزات والأنماط من الصور. في هذه الحالة، تُستخدم طبقة Conv2D

لتقسيم الصورة إلى عدد 14*14 من الرقع باستخدام معلمة الخطوة. تحدد الخطوة الحجم الذي سينزلق بها المرشح

فوق بيانات الإدخال. في هذه الحالة، نظرًا لأن صورنا هي 224×224 وأبعاد الرقعة هي 16×16 فإن حجم الخطوة هو 16.



الشكل (4-4) تقسيم الصور الى رقع [35]

2.4.4 ضبط (fine tuning) نموذج ViT_base_patch_16_224

الضبط في التعلم العميق هو شكل من أشكال نقل التعلم حيث يؤخذ نموذج مدرب مسبقًا، على مجموعة بيانات كبيرة بغرض تنفيذ مهمة عامة مثل التعرف على الصور، ومن ثم إجراء تعديلات طفيفة على بياناته الداخلية [47].

من خلال التدريب المسبق للنموذج، فإنه يتعلم تمثيلًا داخليًا للصور يمكن استخدامه بعد ذلك لاستخراج الميزات من مجموعة البيانات Market. في هذا البحث تم تطوير نسخة مخصصة من نموذج محول الرؤية "vit_base_patch16_224" لمهام التعرف على الأشخاص باستخدام مجموعة بيانات Market-1501. يتم ضبط النموذج عن طريق تعديل معلمات معينة وإضافة طبقات، حيث يتم تعيين عدد فئات هوية الشخص على

751، ويتم تقسيم عملية استخراج الميزات إلى 14 جزءًا باستخدام التجميع المتوسط. يتم تجميد رأس نموذج ViT المدرب مسبقًا، مما يضمن عدم تحديث بعض الأوزان المدربة مسبقًا أثناء التدريب. يتم تطبيق تنظيم التسرب الإضافي (باحتمالية 0.5)، تمكن هذه التعديلات النموذج من التعامل بشكل أكثر فعالية مع مهام إعادة تحديد هوية الشخص من خلال استخراج الميزات ذات الصلة والحد من الإفراط في التدريب. تمت برمجة النموذج باستخدام مكتبة Pytorch.

3.4.4 خوارزمية العمل المقترحة

- اقرأ مجموعة البيانات Market.
 - قم بتغيير حجم الصور إلى 224*224 بكسل.
 - قسم مجموعة البيانات إلى مجموعة تدريب ومجموعة اختبار.
 - قم بتحميل نموذج ViT المدرب مسبقًا باستخدام نقل التعلم.
 - اضبط معلومات النموذج لتتناسب مجموعة البيانات بحيث يكون عدد الفئات 1501، عدد رؤوس التصنيف $z=12$ ، المحسن هو Adam.
 - عرف مصنف لكل رأس تصنيف.
 - درب النموذج بحيث يكون عدد دورات التدريب (epochs=30).
 - لكل صورة في جميع الدفعات (Batches):
1. استخراج التضمينات المحلية والزمنية. (Local and Temporal Embeddings).
 2. أضف أوزان التضمينات (Weigh Embeddings).
 3. تطبيق المشفر (Encoder) للحصول على الرموز العالمية والمحلية (Global and local tokens).
 4. تكوين مصفوفة من الرموز العالمية والمحلية.
 5. تنفيذ التجميع المتوسط (AveragePooling).
 6. تمرير المخرجات إلى المصنف (classifier).
 7. احسب الضبط والخطأ.

8. احسب متوسط الدقة والخطأ لكل دفعة (batch).

- احسب الضبط والخطأ للدورة.
- اطبع الضبط والخطأ.

تم تدريب النموذج باستخدام anaconda بمعالج (intel core i5-12th) وذاكرة (16) جيجا بايت. جاءت

النتائج جيدة يوضح الشكل (4-5) دقة التدريب والاختبار حيث تصل الدقة إلى 98.43% .



الشكل (4-5) دقة التدريب والاختبار الناتجة عن تطبيق نموذج ViT

5.4 مقارنة نتائج تطبيق CNN و ViT ومقارنة نموذج ViT مع بعض الأعمال المنشورة

يوضح الجدول (4-9) الدقة لكل من نمودي CNN و ViT المقترحين

الجدول (4-9) مقارنة بين نمودي CNN و ViT

dataset	CNN	ViT
Market	96.23%	98.43%

يتفوق أداء ViT على CNN وذلك لأن

- CNN تقوم باستخراج الميزات المحلية ، بينما Vision Transformers فهو أكثر مرونة، حيث يلتقط العلاقات العامة والمحلية في الصور.
 - ينمذج ViT بشكل فعال الترابطات طويلة المدى باستخدام آلية الانتباه.
 - إن استخدام تقنية نقل التعلم زادت من قدرة ViT على استخراج الميزات وربطها ببعضها.
- يوضح الجدول (4-10) مقارنة نتائج نموذج ViT مع بعض الأعمال المنشورة.

الجدول (4-10) مقارنة نتائج نموذج ViT مع بعض الأعمال المنشورة

<i>Approach</i>	<i>dataset</i>	<i>technique</i>	<i>advantages</i>	<i>disadvantages</i>	<i>Accuracy</i>
<i>Our method</i>	<i>Market</i>	<i>Transfer_learning</i>	يركز على الميزات المحلية والعالمية	يحتاج إمكانية حسابية عالية	98.43%
<i>Sharma, c et al(2021)</i>	<i>Market</i>	<i>La_transformer</i>	دمج الميزات المحلية والعالمية مع التركيز على الميزات المحلية	بعض الإهمال على الميزات العالمية	98.27%
<i>Hu.B et al(2024)</i>	<i>Market</i>	<i>personVit</i> (<i>self_supervised_learning</i>)	الاعتماد على التعلم المراقب	لا يعتمد على البيانات الموسومة	96.8%

6.4 ملخص نتائج الفصل الرابع

- إن نموذج CNN المقترح بسيط وبنفس الوقت جيد في التعرف على الأشخاص حيث وصلت الدقة إلى 96.23%.
- إن استخدام تكثير البيانات سوف يقلل من حجم الذاكرة المستخدمة مع زيادة في عدد الصور الأمر الذي يقلل الإفراط في التدريب وهذا ما يفسر زيادة الدقة حيث أنه من أجل 200 شخص تم زيادة عدد الصور إلى 21557

صورة.

- على الرغم من أن استخدام طريقة تشبه الخوارزمية الجينية سوف يزيد من زمن المعالجة ولكنه يمكننا من اختيار أفضل قيم للبارامترات.
- إن قدرة ViT على نمذجة العلاقة بين الأجزاء البعيدة من الصورة (على سبيل المثال، بين حذاء الشخص ورأسه) تجعله أكثر قدرة على التعامل مع الأنماط المرئية المعقدة مقارنة بالطرق التلافية البحتة [36].
- بسبب آلية ViT القادرة على تقسيم كل صورة إلى رقع غير متداخلة [35]، فسوف يتم التعامل مع كل رقعة كرمز منفصل في تسلسل الإدخال. ويتبع ذلك فهرسة التضمينات الموضعية، مما يضمن عدم فقد المعلومات المكانية حول مكان كل رقعة في الصورة الأصلية.

7.4 خاتمة الفصل الرابع

لقد نفذنا خوارزميتي CNN و ViT على نفس مجموعة البيانات Market وقارنا النتائج التي حصلنا عليها حيث تبين أن نموذج ViT المقدم أفضل ومن ثم قارنا بين نموذج ViT المطبق وبعض الأعمال المنشورة التي تستخدم تقنية ViT على نفس مجموعة البيانات حيث تفوق نموذجنا المقدم على بقية النماذج المستخدمة في المقالات الأخرى.

الفصل الخامس

تنفيذ شبكات التعلم العميق باستخدام بيانات المشي الناتجة عن

الحساسات

Implementing Deep Learning Networks Using Sensor Gait Data

الفصل الخامس

تنفيذ شبكات التعلم العميق باستخدام بيانات المشي الناتجة عن الحساسات

1.5 مقدمة

تستخدم النماذج المقترحة في هذا الفصل قياسات الاستشعار كميزات للمشبي، بسبب سهولة استخدام أجهزة الاستشعار وتكلفتها المنخفضة، هناك العديد من الدراسات التي استخدمت هذه البيانات كمدخلات ففي [57] ناقش الباحثون التعرف على المشي باستخدام أجهزة استشعار يمكن ارتداؤها وركزوا على تطبيق الشبكات القائمة على التعلم العميق. وبينما أظهرت هذه الشبكات أداءً أفضل مقارنة بالطرق التقليدية، إلا أنها غالباً ما تعاني من تعقيد نموذجي مرتفع، مما يؤدي إلى استهلاك للطاقة كبير وتأخير زمني بالنسبة للأجهزة القابلة للارتداء منخفضة الطاقة.. وفي [58] درس المؤلفون الإمكانية الواعدة لتحليل المشي باستخدام أجهزة الاستشعار القصورية للتعرف على المستخدم حيث اقترحوا نهجاً جديداً لتوسيع بيانات المشي القائمة على أجهزة الاستشعار القصورية للتغلب على تحدي نقص البيانات تم تقديم خوارزميتين لتحويل البيانات مع الحفاظ على قوائم العنونة. وفي [59] كشف المؤلفون عن اعتماد واسع على شبكات التلاف العصبونية (CNNs) أو نماذج الذاكرة القصيرة والطويلة الأمد (LSTM) لاستخراج ميزات المشي. ومع ذلك، لاحظوا قيداً كبيراً في هذه الأساليب فعلى الرغم من أن شبكات CNNs تتفوق في النقاط جميع الخصائص، إلا أنها تواجه صعوبة في الحفاظ على المعلومات الزمنية التي تعد ضرورية لتحليل المشي. وعلى النقيض من ذلك، تواجه نماذج LSTM، التي تتميز بقدرتها على التعامل مع البيانات الزمنية، تحديات أداء مع التسلسلات الطويلة.

2.5 البيئة البرمجية

تم استخدام Google Colaboratory كبيئة برمجية في هذا الفصل، يعتمد على تطبيق jupyter notebook (بيئة تطوير تفاعلية لكتابة الأكواد تتيح واجهته المرنة للمستخدمين تكوين وتنظيم العمل في علوم

البيانات والحوسبة العلمية والتعلم الآلي)، يتم تشغيل Google Colaboratory على السحابة (Cloud) حيث يتيح للمستخدمين تدريب نماذج التعلم الآلي والتعلم العميق على وحدات المعالجة CPU و GPU و TPU.

استخدمنا في هذا الفصل Google Colaboratory بوحدة معالجة T4 Gpu وذواكر 12 جيجا تم رسم المنحنيات باستخدام مكتبة utils.

3.5 تنفيذ شبكة U-Net للتمييز بين فترات المشي والراحة

بما أن الهواتف الذكية تستخدم لجمع البيانات بشكل عشوائي [60] أي أننا لانعرف متى وأين وكيف سيتم استخدام الهواتف الذكية، وبالتالي، تتكون البيانات الملتقطة من فترات المشي وفترات الراحة، ولكن بيانات المشي فقط هي التي تستخدم لاستخراج سمات المشي وتحديد هوية الشخص. وبالتالي، يجب تقسيم سلاسل البيانات إلى فترات مشي وفترات راحة [52].

ونظرًا لأن بيانات المشي وبيانات الراحة مختلفة دلاليًا وأن السلاسل الزمنية الملتقطة مستمرة في كل من المجالين المكاني والزمني، سنقترح بنية مستوحاة من شبكة U-net وتدريبها باستخدام مجموعة البيانات whuGAIT وتحديدًا مجموعتي البيانات 7 و 8.

1.3.5 وصف مجموعتي البيانات 7 و 8

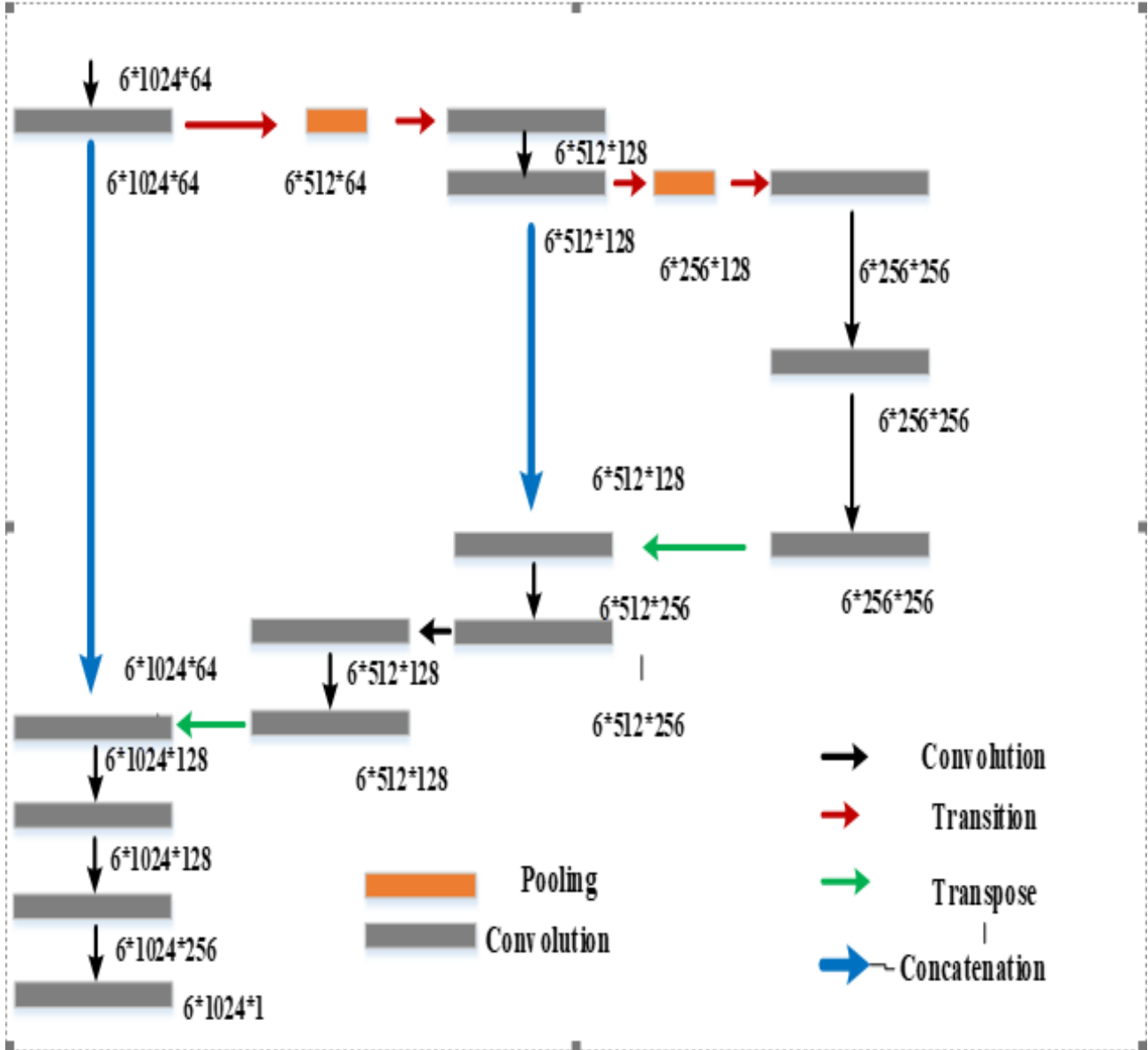
قوائم عناوين المجموعتين 7,8 هي صفر أو واحد؛ 0 يعني عدم المشي و 1 يعني المشي. تم جمع مجموعتي البيانات المقترحة في ظل حالة عدم القيد على المستخدمين دون معرفة متى وأين وكيف يمشي المستخدم.

- **مجموعة البيانات 7:** تحتوي على 577 عينة من 10 أشخاص، من بين هذه العينات، يتم استخدام 519 عينة للتدريب و 58 عينة للاختبار. كل من عينات التدريب والاختبار من نفس الأشخاص العشرة.
- **مجموعة البيانات 8:** تحتوي على 1354 عينة من 118 شخصًا، من بين هذه البيانات، يتم استخدام 1022 عينة من 20 شخصًا للتدريب، ويتم استخدام 332 عينة من 98 شخصًا آخرين للاختبار.

2.3.5 الشبكة المقترحة لفصل عينات المشي والتدريب

تشكل شبكة U-Net الموضحة بالشكل (1-5) الجزء الأهم من البنية المستخدمة بينما يمكن تمثيل

الربط بين الطبقات بالشكل (2-5)، وتتكون آلية عمل U-Net في هذا البحث من المراحل التالية:



الشكل (1-5) بنية U-Net المستخدمة (الباحثة)

- أولاً يتكون الدخل من قياسات الحساسات الستة المأخوذة 1024 مرة بشكل زمني متسلسل.
- في الطبقة (1) يتم اجراء التلاف باستخدام 64 مرشح في كلا المرحلتين فينتج لدينا صورة بحجم

64*1024*6 حيث 64 هو عدد الصور الناتجة ويسمى أيضا عمق الصورة.

- في الطبقة التي تليها نطبق عليها التجميع بحجم نواة مساو ل 1*2 وبالتالي ينخفض حجم الصورة الى النصف فيصبح 64*512*6.
- في الطبقة التي تليها نطبق أيضا التلاف بعدد مرشحات 128 فيصبح حجم الصورة 128*512*6 وهكذا بالنسبة لطبقات التجميع والتلاف.
- بالنسبة لطبقة النقل: يعطى حجم الخرج بالعلاقة (1-5) [61] (حيث طبقنا conv2d_transpose):

$$\text{Out_put} = \text{stride}(\text{input} - 1) + \text{conv_kernel} - 2 * \text{padding} \quad (1-5)$$

$$\text{Out_put} = 512$$

نلاحظ أن هذه الطبقة تقوم بزيادة حجم الصورة بعد أن تم تصغيرها في المسار الانكماشى ويصبح حجم الصورة 128*512*6.

- في طبقة الربط concatenation يتم جمع خرجي الطبقتين المتتاليتين والمتساويتين وينتج الخرج 256*512*6.
- في طبقة transpose الثانية نطبق نفس القانون فيصبح لدينا الحجم 1024.
- نضيف تبطين (padding) في بعض الطبقات حيث يعطى حجم التبطين بالقانون التالي يعطى حجم التبطين بالعلاقة (2-5) [61]:

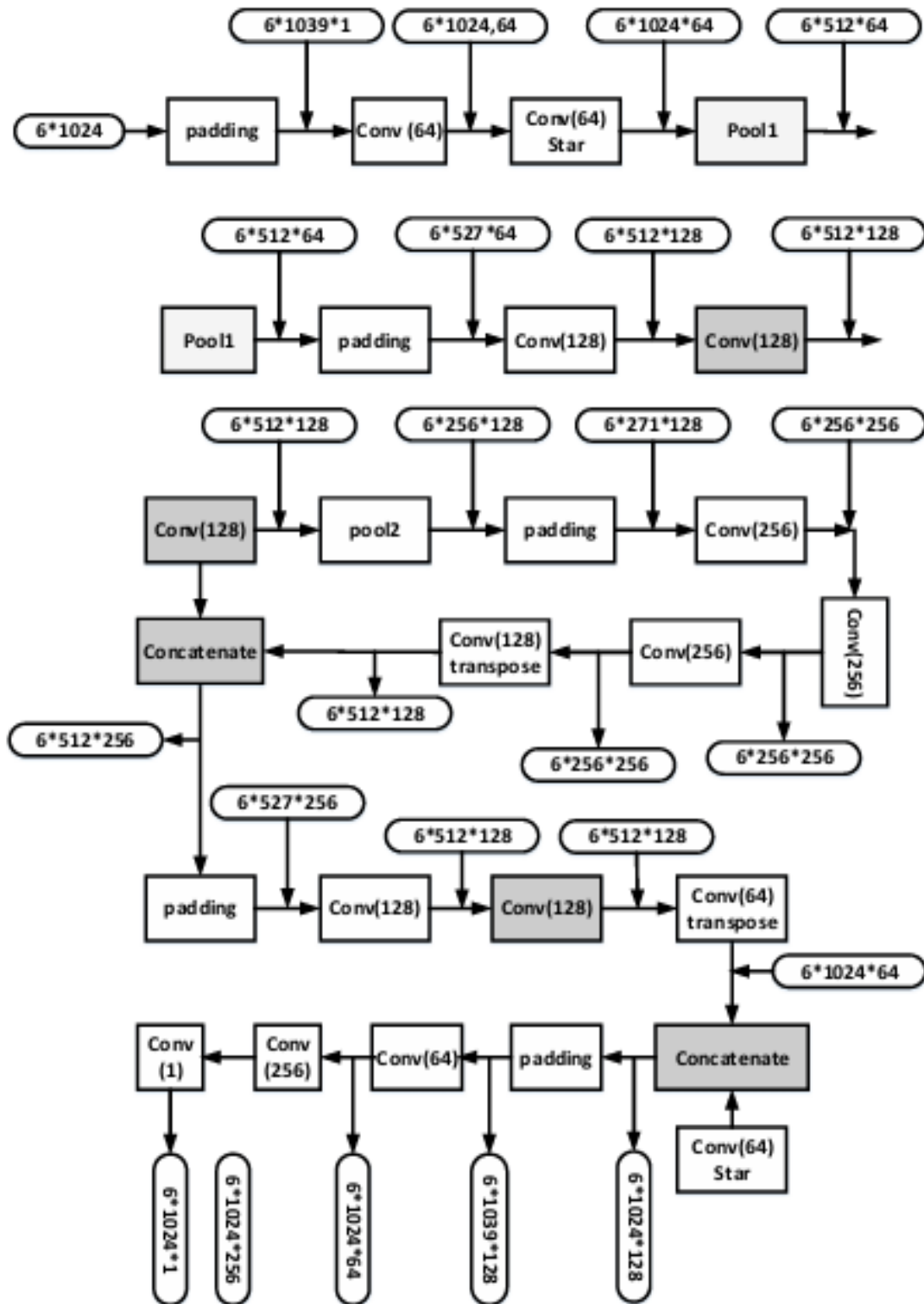
$$\text{Padding} = \frac{k-1}{2} \quad (2-5)$$

$$K=16, 2 * \text{padding}=15$$

وبالتالي يزداد البعد 1024 بهذا المقدار ذلك أننا لا نستطيع زيادة البعد 6 وهذا ما يفسر قيمة البعد 1039. لقد تم تنظيم الطبقات المستخدمة وبارامتراتهما في الجدول (1-5).

الجدول (5-1) الطبقات المستخدمة في فصل فترات الراحة عن المشي وبارامتراتهما

Layer_name	Kernel_size	Kernel_num	Stride	Feature_map
Conv(64)	1*16	64	1*1	6*1024*64
Conv(64)star	1*16	64	1*1	6*1024*64
Pool1	1*2	/	1*2	6*512*64
Conv(128)	1*16	128	1*1	6*512*128
Conv(128)	1*16	128	1*1	6*512*128
Pool2	1*2	/	1*2	6*256*128
Conv(256)	1*16	256	1*1	6*256*256
Conv(256)	1*16	256	1*1	6*256*256
Conv(256)	1*16	256	1*1	6*256*256
Conv(128)transpose	1*2	128	1*2	6*512*128
concatenate	/	/	/	Conv(128)+conv(128)transpose= 6*512*128+6*512*128=6*512*256
Conv(128)	1*16	128	1*1	6*512*128
Conv(128)	1*16	128	1*1	6*512*128
Conv(64)transpose	1*2	64	1*2	6*1024*64
concatenate	/	/	/	Conv(64)star+ Conv(64)transpose= 6*1024*64+6*1024*64=6*1024*128
Conv(64)	1*16	64	1*1	6*1024*64
Conv(256)	1*16	256	1*1	6*1024*256
Conv(1)	1*16	1	1*1	6*1024*1



الشكل (5-2) ربط الطبقات المستخدمة في فصل عينات المشي عن الراحة (الباحثة)

3.3.5 نتائج التصنيف

بالنسبة لمجموعة البيانات 7 فإن بيانات التدريب وبيانات الاختبار مأخوذة على التوالي من نفس الأشخاص؛ حققت طريقتنا دقة تدريب 88.9% ودقة اختبار 86.45% بعد 15 تكرارًا كما هو موضح بالشكل (3-5). أما بالنسبة لمجموعة البيانات 8 فإن بيانات التدريب وبيانات الاختبار مأخوذة على التوالي من أشخاص مختلفين؛ حققت الطريقة المتبعة دقة تدريب 85.3% ودقة اختبار 78.2% بعد 15 تكرارًا. تعتبر النتائج جيدة لأن عدد التكرارات قليل وبالتأكيد ستتحسن الدقة في حال زيادة عدد التكرارات.

```
Step 15, Minibatch Loss= 0.2650, Training Accuracy= 0.894
Step 15, Minibatch Loss= 0.2931, Training Accuracy= 0.869
Step 15, Minibatch Loss= 0.2195, Training Accuracy= 0.928
Step 15, Minibatch Loss= 0.2025, Training Accuracy= 0.924
Step 15, Minibatch Loss= 0.2342, Training Accuracy= 0.915
Step 15, Minibatch Loss= 0.2155, Training Accuracy= 0.917
Step 15, Minibatch Loss= 0.1756, Training Accuracy= 0.930
Step 15, Minibatch Loss= 0.2245, Training Accuracy= 0.923
Step 15, Minibatch Loss= 0.2846, Training Accuracy= 0.890
Step 15, Minibatch Loss= 0.2316, Training Accuracy= 0.910
Step 15, Minibatch Loss= 0.2535, Training Accuracy= 0.898
Step 15, Minibatch Loss= 0.2520, Training Accuracy= 0.889
*****
Testing Accuracy: 0.8645777
*****
Optimization Finished!
```

الشكل (3-5) تدريب النموذج على مجموعة البيانات 7

4.5 تحسين خوارزمية CNN ودمجها مع تقنيات خوارزمية RNN

بعد دراسة طريقة استخراج عينات الحركة والسكون من البيانات الخام في الفصل السابق واختبار دقة النموذج المقترح في فصل تلك العينات عن بعضها، سوف نقترح نموذج تعلم العميق للتعرف على الأشخاص بناءً على خصائص المشية. يستخدم الإطار المقترح نموذج التعلم العميق المدرب على بيانات من ثلاثة جيروسكوبات وثلاثة مقاييس تسارع وذلك لتحديد هوية الأشخاص، حيث تعتبر البيانات الخام للمستشعرات المجدولة ضمن مجموعات البيانات هي بصمة المشية.

1.4.5 وصف مجموعات البيانات المستخدمة

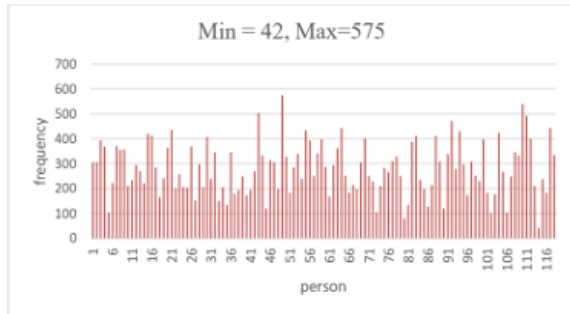
سنستخدم في هذا البحث مجموعات البيانات (1,2,3,4) التابعة لمجموعة البيانات whuGAIT:

- **مجموعة البيانات 1:** تم جمع مجموعة البيانات هذه من 118 موضوعاً حيث جمعت بيانات المشية في خطوات، بناءً على النتائج التي تقيد بأن بيانات الخطوتين لها أداء جيد في التعرف على المشية، جمعت عينات المشية عن طريق تقسيم منحنى المشية إلى خطوتين مستمرتين. وفي الوقت نفسه، قسمت عينة واحدة إلى طول ثابت يبلغ 128 باستخدام دالة الاستيفاء الخطي [52]. من أجل تكبير مقياس مجموعة البيانات، لقد تم تنفيذ تداخل بخطوة واحدة بين عينتين متجاورتين لجميع الموضوعات. وبهذه الطريقة، تم جمع إجمالي عدد 36844 عينة مشية. يتم فرز هذه العينات حسب الوقت المخصص لكل موضوع، نختار أول 90% من العينات للتدريب والباقي 10% للاختبار. يوجد 33104 عينة تدريب و3740 عينة اختبار، بدون تداخل بين المجموعتين الفرعيتين.
- **مجموعة البيانات 2:** تم جمع مجموعة البيانات هذه من 20 موضوعاً. قسم منحنى المشي إلى عينات من خطوتين وتم استيفائها بنفس الطول وهو 128. ونظرًا لأن كل موضوع في مجموعة البيانات هذه يحتوي على كمية أكبر بكثير من البيانات مقارنة بتلك الموجودة في مجموعة البيانات رقم 1، فلم ينفذ التداخل بين العينات، تم جمع إجمالي عدد 49275 عينة، حيث تم استخدام 44339 عينة للتدريب، والباقي 4936 للاختبار.
- **مجموعة البيانات 3:** تم جمع مجموعة البيانات هذه على 118 موضوعاً كما في مجموعة البيانات رقم 1، وخلافًا لمجموعة البيانات رقم 1، يقسم منحنى المشي باستخدام طول زمني ثابت، بدلاً من طول الخطوة، تجمع العينات بفواصل زمنية يبلغ 2056 ثانية. وبينما يبلغ تردد جمع البيانات 50 هرتز، فإن طول كل عينة هو أيضًا 128. يجري تداخل لمدة 1.28 ثانية لتكبير مجموعة البيانات. تم جمع إجمالي عدد 29274 عينة، حيث تم استخدام 26283 عينة للتدريب، والباقي 2991 للاختبار.
- **مجموعة البيانات 4:** تم جمع هذه المجموعة من البيانات من 20 مشاركاً، حيث قسم منحنى المشية على فترة زمنية قدرها 2.56 ثانية، ولم يجري أي تداخل بين العينات. وأخيرًا، تم جمع إجمالي عدد 39314 عينة، حيث

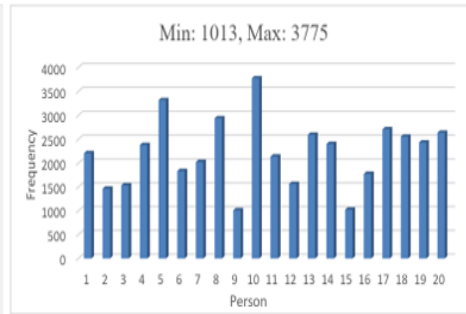
تم استخدام 35373 عينة للتدريب، والباقي 3941 عينة للاختبار .

2.4.5 تحليل مجموعة البيانات

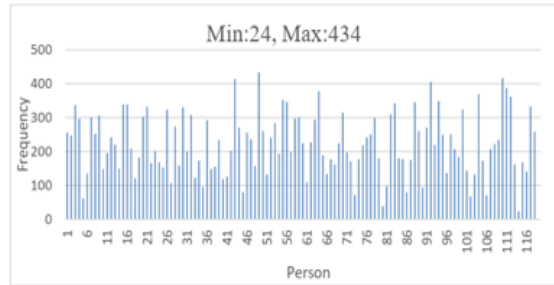
تم تحليل مجموعات البيانات المستخدمة، مع إظهار مخطط الهستوجرام الذي يوضح التردد لكل شخص في مجموعة البيانات لمجموعات البيانات الأربع. تظهر النتائج الموضحة في الشكل (4-5) عدم تجانس في عدد ميزات المشي حيث أن عدد العينات تختلف من شخص إلى آخر وبالتالي من الجيد تطبيق تقنيات مختلفة للحصول على نموذج متين يتعرف على جميع الأشخاص بدقة بالرغم من اختلاف عدد عيناتهم يبين الجدول (2-5) أكبر



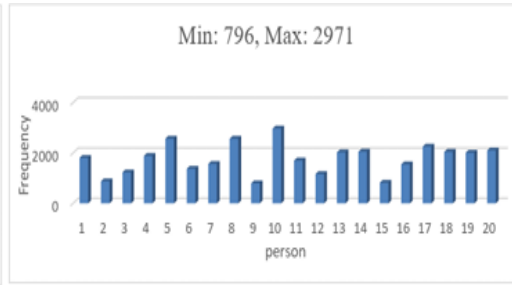
1



2



3



4

الشكل (4-5) الهستوجرام لمجموعات البيانات 4,3,2,1

قيمة وأصغر قيمة لعدد العينات في كل من مجموعات البيانات.

الجدول (2-5) أكبر قيمة وأصغر قيمة لعدد العينات في مجموعات البيانات 1,2,3,4

Dataset	Frequencies	
	Min	Max

Dataset #1	42	575
Dataset #2	1013	3775
Dataset #3	24	434
Dataset #4	796	2971

3.4.5 اختبار الشبكة العصبونية التلافية CNN مع نوعين من توابع التفعيل

إن وحدة التصحيح الخطية ReLU (Rectified Linear Unit) و وحدة التنشيط الخطية الأسية (Exponential Linear Unit) ELU هما دالتا تنشيط تستخدمان بشكل شائع في نماذج التعلم العميق.

يتم تعريف ReLU [62] بالعلاقة الرياضية $f(x) = \max(0, x)$ ، حيث تقوم بإخراج قيمة الإدخال إذا كانت موجبة، وصفرًا بخلاف ذلك. تعتبر ReLU فعّالة من الناحية الحسابية، كما أنها تساعد في التخفيف من مشكلة تلاشي المشتقات (vanishing gradient)، ومع ذلك، تعاني ReLU من مشكلة "موت ReLU"، حيث يمكن أن تصبح الخلايا العصبونية غير نشطة وتعطي الخرج صفرًا مهما كانت المدخلات. يمكن أن يحدث هذا أثناء التدريب، وبمجرد أن تصبح الخلية العصبونية غير نشطة، فإنها لم تعد قادرة على تحديث أوزانها، مما يؤدي إلى خلايا عصبية ميتة لا تساهم في التعلم [63].

يتم تعريف ELU على أنها $f(x) = x$ إذا كانت $x > 0$ ، و $\alpha * (\exp(x) - 1)$ إذا كانت $x \leq 0$ ، حيث α هو ثابت موجب صغير.

تعالج ELU مشكلة "موت ReLU" (dying ReLU) من خلال السماح للقيم السالبة بتنشيط الخلايا العصبونية [64]، مما يساعد في الحفاظ على الخلايا العصبونية حية ونشطة أثناء التدريب [58]. توفر ELU أيضًا السلاسة والاستمرارية لكل من المدخلات الموجبة والسالبة.

يوضح الجدول (3-5) المقارنة بين ReLU و ELU.

الجدول (3-5) المقارنة بين ReLU و ELU

الميزات	ELU	ReLU
نطاق التنشيط	$[-\alpha, \infty]$	$[0, \infty]$
الاستمرارية	خطي متقطع ولكنه ناعم وسلس ومستمر للدخل الإيجابي والسلبى	خطي متقطع باستثناء الصفر يملك زاوية حادة
موت العصبون	أكثر متانة وفعالية في التعامل مع موت العصبون	قد يؤدي الى موت العصبون
الفعالية الحسابية	أكثر تعقيدا في الحساب	سهولة الحساب وأسرع في الحساب
تلاشي المشتقات	أكثر عرضة لمشكلة تلاشي المشتقات	غير مشبعة للدخل الإيجابي مما يخفف من مشكلة تلاشي المشتقات

في الممارسة العملية، يعتمد اختيار حجم الدفعة على عوامل مختلفة، بما في ذلك موارد الأجهزة المتاحة وحجم مجموعة البيانات وتعقيد النموذج ومشكلة التعلم المحددة. من الشائع تجربة أحجام دفعات مختلفة ومراقبة تأثيرها على ديناميكيات التدريب وأداء النموذج في مجموعة التحقق لتحديد حجم الدفعة الأمثل لمهمة معينة.

تم اعتماد ReLU بحجم دفعة أقل، بعد العديد من التجارب، إن التحسين المطبق على خوارزمية CNN له نتائج جيدة للغاية، حيث وصلت الدقة لقيمة أكبر من 97% تم تلخيص النتائج في الجدول (5-4) الذي يوضح ملخص نتائج تنفيذ CNN مع حالات مختلفة من دالة التنشيط وأحجام الدفعات على مجموعة البيانات 2.

إن الدقة التي حصلنا عليها جيدة ولكن جمع CNN مع RNN وتقنياتها المطورة سوف يعطي نتائج أفضل كما سنوضح بالفقرات التالية لقد تم الجمع بين CNN ومجموعة من خوارزميات RNN والمقارنة بين النتائج.

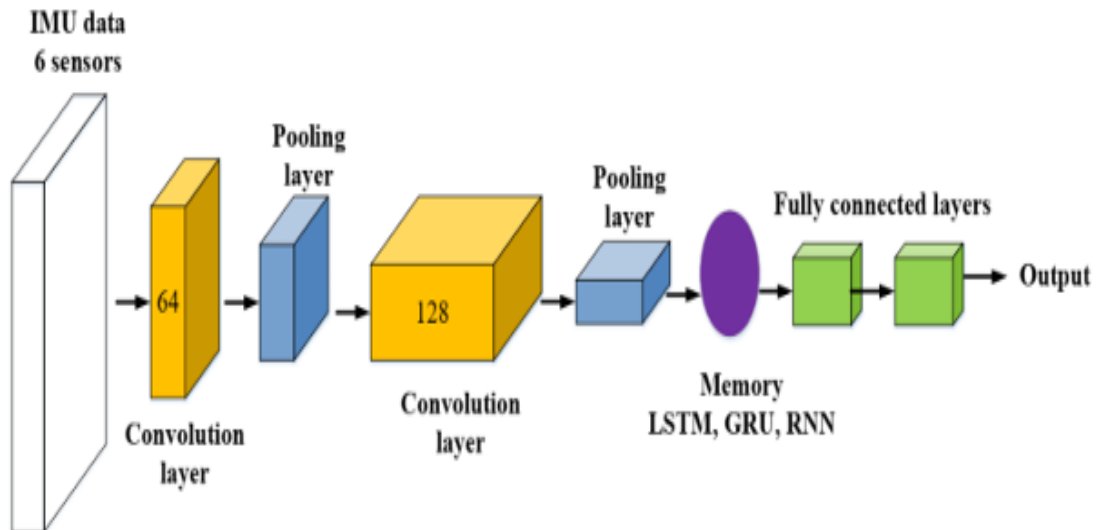
نتائج تنفيذ CNN مع توابيع تفعيل وأحجام دفعات مختلفة

الخطأ		الدقة		الحالة المدروسة
الاختبار	التدريب	الاختبار	التدريب	
0.76131	4.5×10^{-7}	93.69%	1 %	Relu+512
0.6303	0	94,84%	1%	Relu+256
0.76229	0	93.467%	1%	Elu+512

4.4.5 بنية نموذج CNN مع RNN المقترح

يوضح الشكل (4-5) البنية المقترحة لنموذج CNN مع RNN حيث يمكن استخدام وحدات RNN

بأنواعها المختلفة مكان طبقة الذاكرة الموجودة في النموذج.



الشكل (4-5) البنية المقترحة لنموذج CNN مع RNN (الباحثة)

لقد تم الجمع بين CNN و RNN البسيطة ومن ثم تم الجمع مع LSTM مع وبدون طبقات استنظام

الدفة بعدها تمت المقارنة بين النتائج يوضح الجدول (5-5) بنية طبقات CNN والجدول (6-5) طبقات RNN البسيطة و LSTM، نقترح تسمية النموذج بنظام التعرف على المشي باستخدام بيانات وحدة القياس العطالية IMU-GR.

الجدول (5-5) بنية طبقات CNN في نظام IMU-GR

Layer name	Convolution filter number	Kernel_size	Pooling	convolution filter number	Kernel_size	pooling
	64	3	2	128	3	2

الجدول (6-5) طبقات RNN البسيطة و LSTM

Layer name	Number of units
Simple RNN	128
LSTM first senario	64
LSTM second senario	128

تم تصميم النماذج باستخدام مكتبي Tensorflow و Keras وبعد التدريب والاختبار لخصت النتائج بالجدول (7-5):

الجدول (7-5) نتائج تنفيذ نظام IMU-GR

التقنية المستخدمة	CNN+RNN	CNN+LSTM(64)	CNN+LSTM(128)	CNN+LSTM(128)with batch_Normalization

مجموعة البيانات 1	92.54 %	97.64%	97.64%	97.67%
مجموعة البيانات 2	98.34 %	99.28%	99.39%	99.38%
مجموعة البيانات 3	65.33 %	97.18%	97.53%	96.98%
مجموعة البيانات 4	97.22 %	99.32%	99.38%	99.36%

5.5 تحسين CNN ودمجها مع LSTM للتعرف على الأشخاص باستخدام عينات المشي مختلفة الطول

بعد مقارنة عدة تقنيات مختلفة للتعرف على الأشخاص نقترح تقديم نموذج تعلم عميق للتعرف على

الأشخاص اعتمادا على مجموعة بيانات يختلف طول العينات فيها من شخص لآخر .

1.5.5 تحسين مجموعة البيانات OU-ISIR

تم اختيار الجزء (AutomaticExtractionData_IMUZCenter) من مجموعة البيانات OU-ISIR

الذي يحتوي على بيانات مرتبة بشكل عشوائي يوضح الشكل (5-5) ترتيب عينات البيانات الناتجة عن قياسات

السرعة الزاوية والتسارع حيث يمثل N طول العينات ضمن الملف والذي يختلف من ملف لآخر ضمن الجزء

المختار و كل ملف يمثل عينات شخص حيث يوجد ملفين لكل شخص ضمن

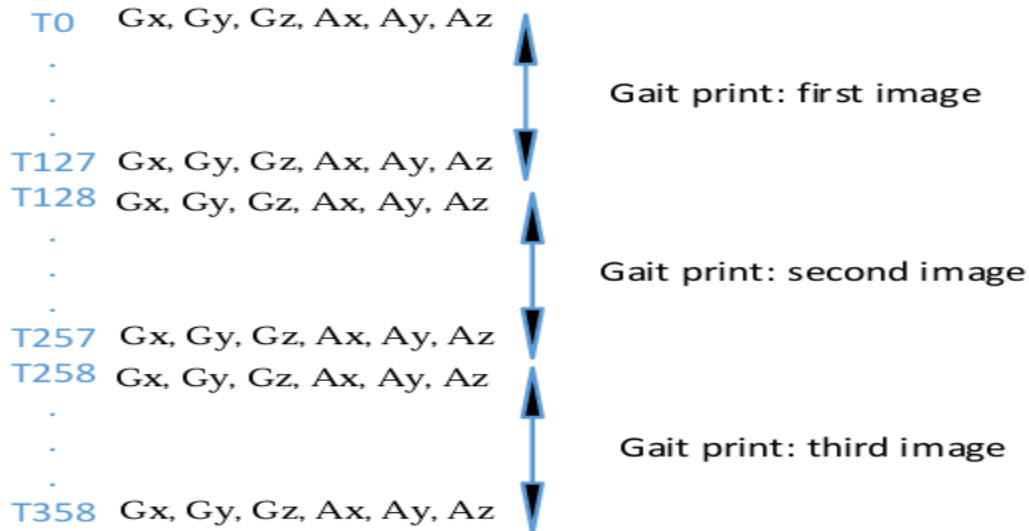
(AutomaticExtractionData_IMUZCenter)، والان سوف نقوم بتقسيم العينات إلى خطوات مجموعة صور

متساوية الأبعاد .

لنفترض أن عدد العينات للشخص "X" في ملفه الأول هو "n"، بالتالي يحتوي الملف على بيانات بحجم $6 * n$ سيتم إعادة تشكيل هذه البيانات لتصبح $6 * n_1 * n_2$ ، حيث يمثل "n₁" الوقت الكافي للنقاط بصمة المشي للشخص، وبالتالي يكون "n₂" هو عدد بصمات المشي لهذا الشخص، وكل بصمة مشي تحتوي على n₁ من بيانات العينات. مثلاً: إذا كان الملف يحتوي على 400 صف فإن n₁=128 تبقى ثابتة و n₂=3 يوضح الشكل (5-6) بنية العينات الجديدة.

T0 Gx, Gy, Gz, Ax, Ay, Az
T1 Gx, Gy, Gz, Ax, Ay, Az
T2 Gx, Gy, Gz, Ax, Ay, Az
T3 Gx, Gy, Gz, Ax, Ay, Az
.
.
.
TN Gx, Gy, Gz, Ax, Ay, Az

الشكل (5-5) ترتيب عينات بيانات OU-ISIR قبل التعديل (الباحثة)



الشكل (5-6) ترتيب عينات بيانات OU-ISIR بعد التعديل (الباحثة)

تم تقسيم مجموعة البيانات إلى جزأين، جزء للتدريب وآخر للاختبار. تم التقسيم يدوياً لضمان احتواء كلا الجزأين على عينات من جميع هويات الأشخاص الموجودة في مجموعة البيانات. تم اختيار عدد العينات الذي يعتبر كافياً لحفظ بصمة المشي ليكون 128 عينة، وذلك بعد العديد من المحاولات والتجريب في اختيار طول بصمة المشي.

2.5.5 هيكل النموذج المقترح للتعرف على الأشخاص باستخدام عينات مختلفة الطول

يوضح الشكل (5-7) بنية النموذج وأبعاد الدخل والخرج لكل طبقة، حيث استخدمنا المعالجة على التوازي يوجد طبقة Masking لإخفاء الميزات الغير مهمة. يتم استخدام طبقة LSTM لالتقاط التسلسل الزمني للبيانات، تليها طبقة Dropout للتقليل من فرط التكيف. بالإضافة إلى ذلك، يتم استخدام سلسلة من الطبقات التلافية مع طبقات التقييس الدفعية وتوابع التفعيل لالتقاط الأنماط المحلية (local patterns) في البيانات. كما يتضمن النموذج مكونات مخصصة وهي كتل Squeeze and Excite لتعزيز إعادة تقويم الميزات. يتم تحديد معاملات النموذج الفائقة، بما في ذلك عدد وحدات LSTM، ومعدل الإسقاط dropout rate، وأحجام المرشحات، وخطوات الطبقات strides، ومعدل التعلم، ضمن تعريف النموذج. يتم استخدام Keras Tuner لضبط بعض المعلمات الفائقة باستخدام تقنية Random Search، مع تنفيذ متوازي لتسريع العملية حيث يتحقق التنفيذ المتوازي من خلال تنفيذ طبقات CNN و LSTM على التوازي والجمع يتم باستخدام طبقة Concatenating [65]. تُستخدم طريقة tuner. search لبدء عملية ضبط المعلمات الفائقة، بما في ذلك إيقاف التدريب المبكر لمنع فرط التدريب overfitting أثناء التدريب.

1.2.5.5 المكونات المخصصة squeeze and Excite: مهمة هذه الطبقة استخراج شعاع أوزان من الدخل وإعادة ضربه بالدخل وبالتالي النقاط أهم الميزات التي تعبر عن بصمة المشي [66]، تتألف من الطبقات التالية كما هو موضح بالشكل (5-8).

(سيتم شرح الحصول على أبعاد أول كتلة ضمن النموذج أثناء شرح الطبقات):

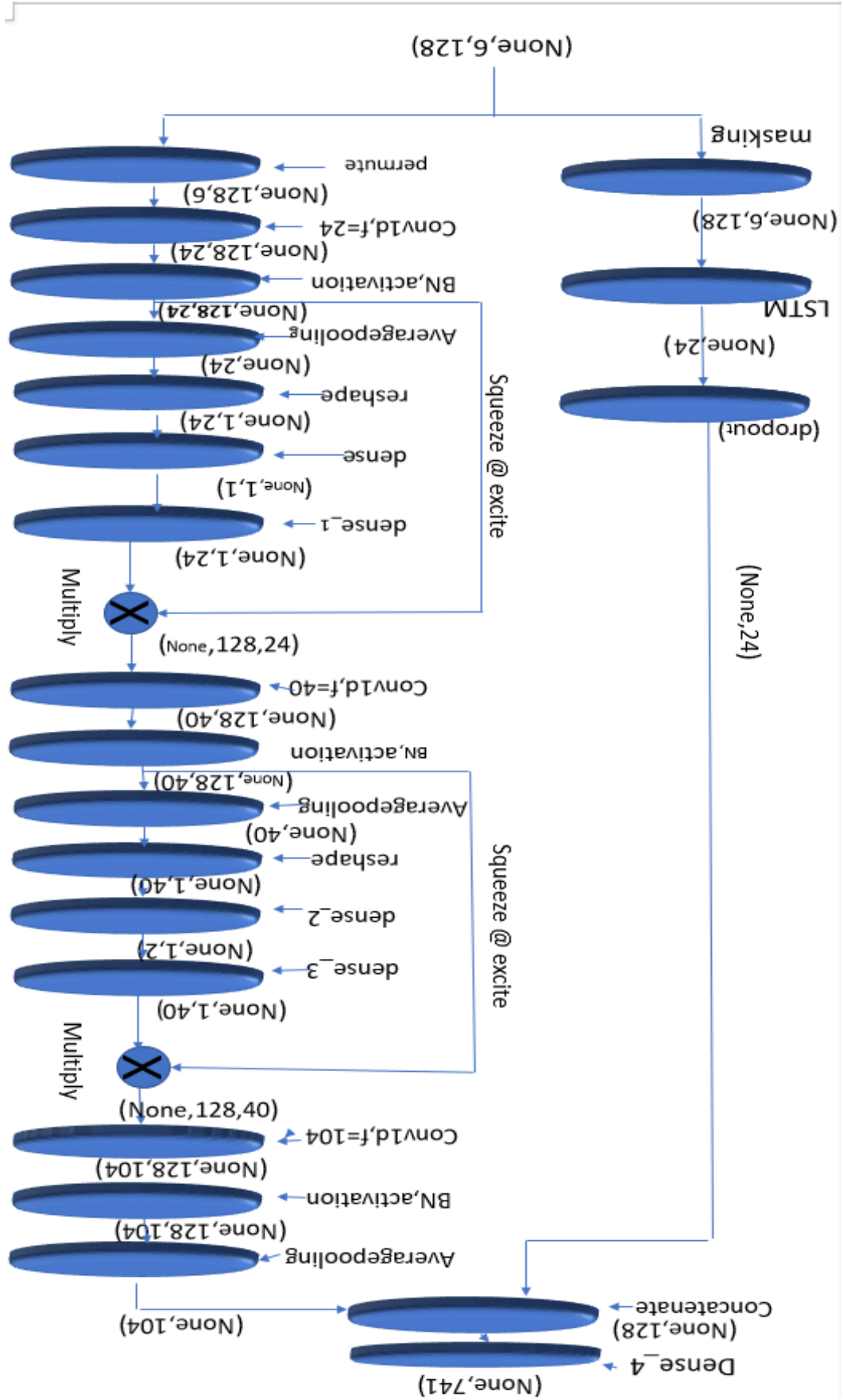
- **طبقة global_average_pooling:** مهمة هذه الطبقة تقليل الأبعاد فإذا كان الدخل (T,C) فإن الخرج يكون (1,C) تنفذ هذه الطبقة مهمتها على مرحلتين المرحلة الأولى خرج طبقة التجميع المتوسط، بما أن الدخل (None, 128, 24) فإن الخرج يكون بالأبعاد (None, 24) بما أن الأبعاد التي نريدها هي (1 * c) حيث c=24 نقوم بعملية reshape فتصبح الأبعاد (None, 1, 24).

- **طبقة fully connected_1:** تقوم هذه الطبقة بتطبيق تابع التفعيل sigmoid الذي يعطي قيما بين الصفر والواحد وذلك ليتكون شعاع الأوزان بأبعاد (none, 1, 24) حيث 24 هو عدد المرشحات (units) المستخدمة ويساوي قيمة أول بارامتر في الطبقة dense [63].

- **طبقة multiply:** التي تقوم بتطبيق الضرب بعنصر (multiplied_element_wise) حيث تكون المداخل (none, 128, 24) و (none, 1, 24) فيصبح الخرج بعد الضرب بالأوزان (none, 128, 24).

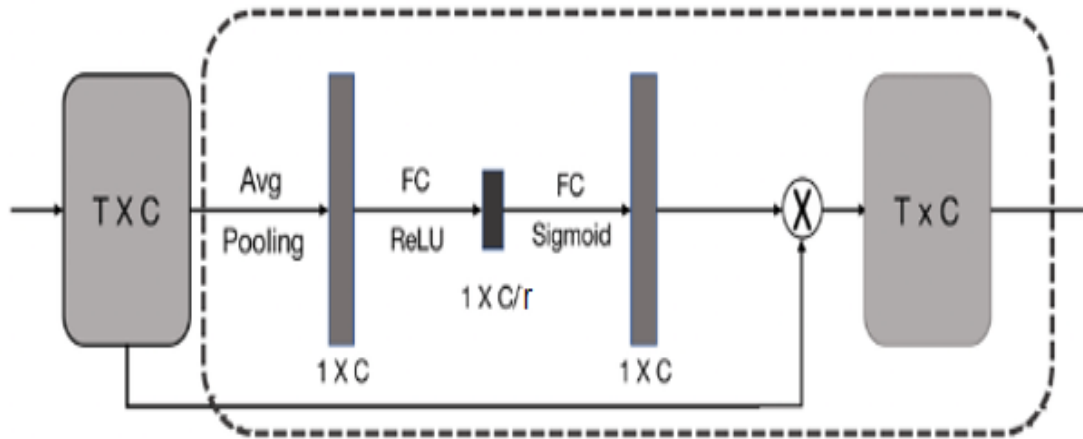
2.2.5.5 عنوان وتقسيم مجموعة البيانات: إن مجموعة البيانات المستخدمة غير معنونة ولكن يمكن الاستفادة من اسم الملف في العنوان طالما يوجد لكل شخص ملفين من ضمن ملفات مجموعة البيانات وبما أن المجموعة غير مقسمة فيجب تقسيم مجموعة البيانات إلى مجموعتي بيانات تدريب واختبار، الشكل (5-9) يوضح الخوارزمية المتبعة والشكل (5-10) يوضح المخطط الصندوقي المتبع لتقسيم مجموعة البيانات.

نقوم بفصل الجزء الرقمي عن الأجزاء النصية للعنوان أما فيما يتعلق بفصل عينات التدريب عن الاختبار نقسم مجموعة البيانات حيث نحدد عدد الصور في كل ملف فإذا كان عدد الصور أصغر من 5 فإن عدد صور الاختبار يساوي 1، إذا كان عدد الصور أصغر من 11 فإن عدد عينات الاختبار يساوي 2 وإذا كان عدد الصور أصغر من 16 فإن عدد عينات الاختبار يساوي 3، طالما أن عدد العينات لم يتجاوز عدد عينات الملف. أخيراً إذا كان عدد الصور أصغر من عدد صور عينات الاختبار، نخزن الصور في مصفوفة الاختبار ونخزن العناوين في قائمة عناوين الاختبار وإلا نخزن الصور في مصفوفة التدريب ونخزن العناوين في قائمة عناوين التدريب.



الشكل (5-7) بنية النموذج المستخدم لاختبار عينات مختلفة الطول (الباحثة)

KerasTuner 3.2.5.5: هو إطار عمل سهل الاستخدام وقابل للتطوير لتحسين المعلمات الفائقة. يمكن تكوين مساحة البحث باستخدام صيغة تعريف التشغيل، ثم الاستفادة من إحدى خوارزميات البحث المتاحة للعثور على أفضل قيم للبارامترات. يأتي KerasTuner مزودًا بخوارزميات التحسين البايزي والبحث العشوائي المضمنة [67] كما تم تصميمه ليكون سهلاً على الباحثين لتوسيعه من أجل تجربة خوارزميات بحث جديدة حيث يتم استدعاؤه كتابع باستخدام وسيط (hp) لتحديد المعلمات الفائقة أثناء إنشاء النموذج. وتحديد مجموعة من القيم المعتاد استخدامها باستخدام تعليمة (hp.Choice) وتميرها إلى الطبقة المراد اختيار بارامترات فتختار خوارزمية البحث أفضل البارامترات بعد تجربة القيم، يوضح الجدول (5-8) البارامترات التي تم اختيارها والقيم المضبوطة باستخدام Keras_Tuner المقابلة، باستخدام Keras Tuner، يصبح ضبط المعلمات الفائقة عملية قابلة للتكرار، مما يسمح بتحقيق نتائج متسقة عبر تجارب مختلفة.



الشكل (5-8) طبقة Squeeze and Excite [66]

سنطلق على النموذج optimized multi LSTM fully Convolution network (OMLSTM-FCN) اسم الشبكة العصبونية المحسنة متعددة المقاييس. يتم تدريب واختبار النموذج عبر google_colab حيث وصلت دقة التعرف إلى 98.23 % على البيانات غير المرئية.

Input: a sequence of structure person's gait data stored in csv files

Output: Data for testing and training and their labels

Start Procedure

Import model libraries

Phase1:

Foreach csv_file

split file name according to the second (-) it contains
store the number in label array

end

phase2:

foreach csv_file in sequence

count the line in the csv_file
calculate the number of images N by dividing the number of lines by 128
if $N < 5$ then the number of test Images $NT=1$
if $N < 11$ then the number of test Images $NT=2$
if $N < 16$ then the number of test Images $NT=3$

foreach row in csv file

if $N < NT$ then store the Images in test array
store the corresponding label in test label array
if $N > NT$ then store the Images in train array
store the corresponding label in train label

end

end

End Procedure

الشكل (5-9) الخوارزمية المتبعة في عنونة وتقسيم الداتا (الباحثة)

الجدول (5-8) قيم البارامترات المضبوطة باستخدام keras Tuner

LSTM number_layer	128
DropOut	0.1
First conv layer	24
First_kernel_stride	3
Second conv layer	40
Second_kernel_stride	3
Third conv layer	104
Third_kernel_size	3

6.5 ملخص نتائج تنفيذ خوارزميات التعلم العميق باستخدام بيانات الحساسات**1.6.5 مناقشة تنفيذ شبكة U-Net باستخدام مجموعتي البيانات (7,8) التابعتان لمجموعة بيانات****whuGAIT**

يعتبر التصميم مناسب لأداء مهمة الفصل بين فترات الراحة والمشي للأسباب التالية:

- استخراج السمات الهرمية: تبدأ بنية الشبكة بطبقات التلاف البسيطة ثم تتعمق تدريجيًا. تحتوي الطبقات في

البداية على مرشحات أصغر (64، 128 مرشحًا)، وهذا من شأنه أن يلتقط السمات الأساسية مثل الحواف

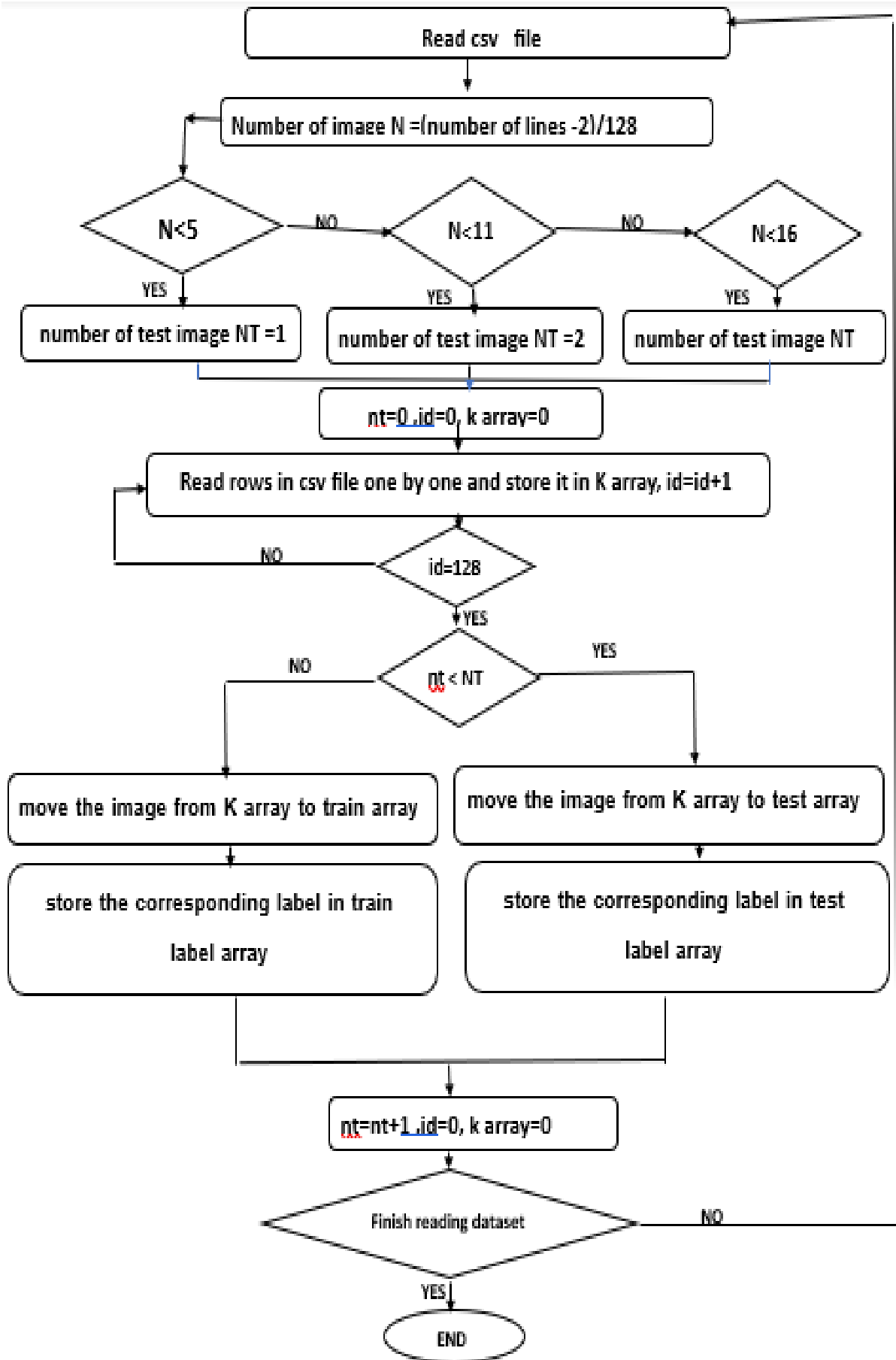
والخطوط العريضة وأنماط الحركة البسيطة للمدخلات ذات الأبعاد الثنائية. ثم تتعمق الشبكة مما يلتقط سمات

أكثر تعقيدًا تتعلق بديناميكيات المشي، والتي تعد بالغة الأهمية للتمييز بين المشي والأنشطة الأخرى.

شبكة مبسطة باستخدام طبقات التجميع: تعمل طبقات التجميع على تبسيط البيانات عن طريق تقليل حجمها

والاحتفاظ فقط بالميزات الأكثر أهمية. وهذا لا يجعل النموذج أسرع وأكثر كفاءة فحسب، بل يساعده أيضًا

على التعامل مع الاختلافات الصغيرة، مثل التغييرات في سرعة المشي أو طول الخطوة.



الشكل (5-10) المخطط الصندوقي لعنونة وتقسيم البيانات (الباحثة)

- **المحافظة على البيانات:** إن وجود طبقة الربط يسمح بتدفق التدرجات بحرية أكبر عبر النموذج، مما يخفف من مشكلة اختفاء التدرجات كما إنها تسمح بإضافة الميزات من جانب المشفر في الشبكة إلى جانب فك التشفير، وهذا يؤدي إلى عدم فقدان المعلومات التي قد تضيع بسبب عملية التجميع على جانب المشفر في الشبكة. كما يسمح للشبكة بالحفاظ على تمثيل أكثر ثراءً لنمط المشية من خلال دمج الميزات المبكرة البسيطة والميزات الأكثر تعقيداً وعمقاً.
- **زيادة عدد المرشحات:** مع تعمق الشبكة، تبدأ بـ 64 مرشحاً وتزداد إلى 256، ثم تنقص مرة أخرى إلى 128، وأخيراً إلى مرشح واحد. يساعد هذا التقدم التدريجي للشبكة على تعلم مجموعة واسعة من الميزات على مستويات مختلفة، وفي النهاية تضيق إلى أهم طبقة وهي الطبقة الأخيرة المسؤولة عن التعرف على ما إذا كان شخص ما يمشي أم لا.

2.6.5 مناقشة اختبار CNN على حجوم دفعات مختلفة وتوابع تفعيل مختلفة

تم اختبار الحالات التالية:

- CNN مع تابع تفعيل RELU وحجم الدفعة (512).
 - شبكة CNN مع تابع تفعيل RELU وحجم الدفعة (256).
 - شبكة CNN مع تابع تفعيل ELU وحجم الدفعة (512).
- إن الخسارة النهائية للاختبار أقل من 1: وهذا يعني أنه بعد تدريب النموذج، تكون الخسارة أقل من 1، من المهم ملاحظة أنه قد لا تكون دالة الخسارة المستخدمة محصورة بالضرورة بين 0 و1، وخاصة في السيناريوهات التي تتضمن نماذج معقدة أو مجموعات بيانات كبيرة. بالإضافة إلى ذلك، قد تشير التقلبات في قيم الخسارة التي تتجاوز 1 إلى محاولة النموذج على التكيف مع البيانات، حيث قد تشير حالات معينة من الخسارة العالية إلى التعديلات الضرورية لتحسين أداء النموذج. علاوة على ذلك، يمكن أن يؤثر اختيار المحسن ومعدل التعلم وحجم الدفعة أيضاً على حجم قيم الخسارة الملحوظة أثناء التدريب. لذلك، في حين أن قيم الخسارة التي تتجاوز 1 قد

تبدو غير تقليدية، إلا أنها يمكن أن تكون مؤشراً على عملية التعلم الديناميكية وقدرة النموذج على التكيف مع تعقيد البيانات. تشير الخسارة النهائية الأصغر إلى أن النموذج تعلم تقديم تنبؤات أفضل، وتم تقليص الفرق بين تنبؤاته والقيم الحقيقية. وهذه نتيجة إيجابية لأنها تشير إلى تحسن أداء النموذج. يبين الشكل (5-11) الاختبار مع حجم دفعات مختلفة بالنسبة لتابعي التفعيل Elu و ReLU.

خلال عملية التدريب، وصلت الخسارة إلى قيم أكبر من 1 وذلك طبيعي فمن الشائع جدًا ملاحظة قيم خسارة أكبر من 1 خلال المراحل المبكرة من التدريب.



الشكل (5-11) خسارة الاختبار مع حجمي دفعة 512 و 256 وتابعي التفعيل Elu, ReLU

خلال الدورات الأولى، يتم تهيئة أوزان النموذج بشكل عشوائي، وقد تكون التوقعات بعيدة عن الأهداف الصحيحة، مع تقدم التدريب، يضبط النموذج أوزانه تدريجيًا من خلال خوارزمية الانتشار العكسي والانحدار التدريجي لتقليل

دالة الخسارة.

3.6.5 مناقشة نتائج تنفيذ نظام IMU-GR باستخدام مجموعات البيانات (1,2,3,4) من مجموعة

البيانات whuGAIT

1.3.6.5 مقارنة بين CNN و CNN+LSTM:

أظهرت النتائج أن CNN+LSTM أفضل من CNN فقط، والأسباب هي:

- **نمذجة بيانات المشية الزمنية:** تتكون بيانات IMU من قياسات سلسلة زمنية من الجيروسكوبات ومقاييس التسارع، والتي تلتقط أنماط حركة الشخص. تعتبر CNN+LSTM مناسبة تمامًا للتعامل مع مثل هذه البيانات المتسلسلة حيث يمكنها التقاط الديناميكيات والأنماط الزمنية الموجودة في حركات الشخص بمرور الوقت بشكل فعال. يتيح مكون LSTM للنموذج التعلم من تسلسل قراءات IMU وفهم كيفية تطور الحركات وتغيرها.
- **المعلومات السابقة:** في مهام التعرف على الشخص، يعد سياق الحركات أمرًا بالغ الأهمية للتحديد الدقيق حيث تسمح اتصالات LSTM المتكررة للنموذج بالحفاظ على الحالات الداخلية، والتي يمكن أن تمثل معلومات سابقة حول حركة الشخص. يساعد هذا الفهم السياقي للنموذج على اتخاذ قرارات أكثر دقة حول هوية الشخص بناءً على أنماط حركته.
- **استخراج السمات والتعلم الهرمي:** تعتبر شبكات CNN فعالة في تعلم السمات الهرمية من البيانات الخام، مثل الأنماط المكانية في الصور. من خلال الجمع بين شبكات CNN وذاكرة LSTM، يمكن للنموذج تعلم التمثيلات الهرمية في كل من المجال المكاني (عبر شبكة CNN) والمجال الزمني (عبر ذاكرة LSTM) [68] يتيح هذا للشبكة التقاط السمات ذات الصلة من بيانات IMU وفهم كل من الترتيب المكاني للمستشعرات والأنماط الزمنية للحركات.
- **تحسين أداء التعرف:** يسمح إدراج ذاكرة LSTM في البنية LSTM على التقاط تطور الحركات بمرور الوقت أمرًا بالغ الأهمية للتمييز الدقيق بين الأفراد المختلفين.

- **تقليل معالجة البيانات المسبقة:** عند استخدام شبكة CNN وحدها، قد يلزم تحويل بيانات ذاكرة IMU إلى تمثيلات ذات حجم ثابت (على سبيل المثال، مخططات طيفية) قبل تغذيتها إلى شبكة CNN [69]، ومع ذلك، باستخدام CNN+LSTM، يمكن استخدام بيانات IMU المتسلسلة الأصلية بشكل مباشر، مما يقلل الحاجة إلى خطوات المعالجة المسبقة وفقدان المعلومات المحتمل.

2.3.6.5 مقارنة بين CNN+LSTM و CNN+RNN

أظهرت النتائج أن CNN+LSTM أفضل، والأسباب هي:

- **معالجة البيانات المرتبطة زمنياً خلال فترة طويلة المدى:** تم تصميم LSTM للتعامل مع التبعيات طويلة المدى، مما يجعلها أكثر ملاءمة للمهام حيث يكون السياق عبر تسلسلات أطول أمراً بالغ الأهمية [70]. وهذا مهم بشكل خاص في معالجة اللغة الطبيعية، والتنبؤ بالسلاسل الزمنية، والمهام الأخرى ذات التبعيات الممتدة. SimpleRNN محدودة في النقاط التبعيات طويلة المدى بسبب مشكلة تلاشي المشتقات، والتي يمكن أن تعيق أدائها في المهام التي تتطلب نمذجة علاقات معقدة على مدى فترات ممتدة أكثر عرضة للإفراط في التدريب وخاصة في المهام المتسلسلة المعقدة. وكما هو معروف فإن عدم التوازن في مجموعة البيانات سيؤدي إلى الإفراط في التدريب. سيعالج استخدام LSTM هذه المشكلة.

المرونة في أنواع البيانات: CNN+LSTM أكثر تنوعاً ويمكن تطبيقها على أنواع مختلفة من البيانات المتسلسلة،

مثل الصوت والفيديو والنص الطبيعي وبيانات السلاسل الزمنية، حيث تكون العلاقات المكانية والزمانية مهمة. في

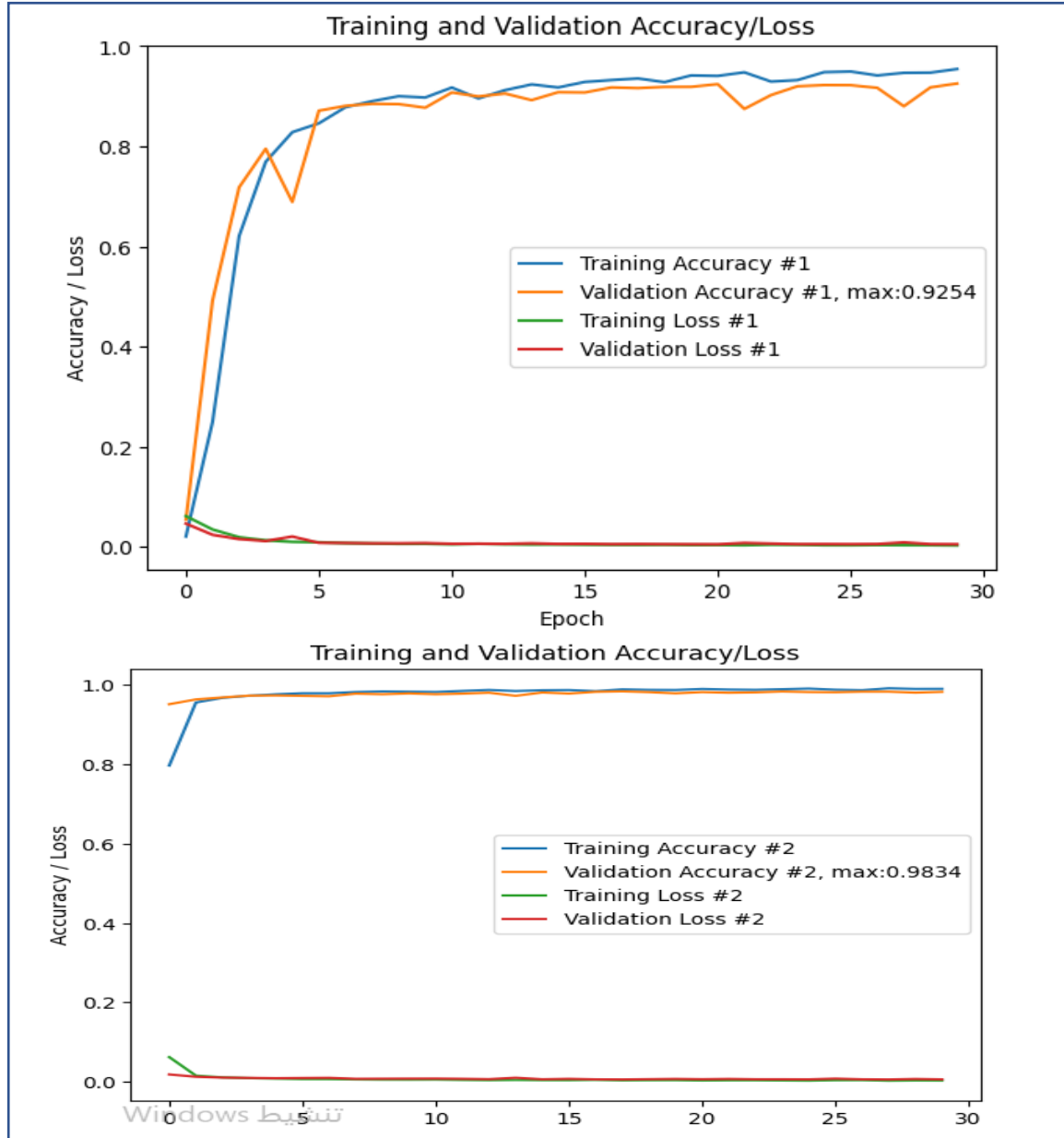
حين أن CNN+SimpleRNN مناسب لمهام متسلسلة معينة، فإن قيوده في التعامل مع التبعيات طويلة الأجل

تحد من إمكانية تطبيقه في سيناريوهات أكثر تعقيداً وواقعية. يوضح الشكل (5-12) الدقة والخطأ عند تطبيق بنية

النموذج CNN+SimpleRNN بالنسبة لمجموعات البيانات 1,2 والشكل (5-13) الدقة والخطأ عند تطبيق بنية

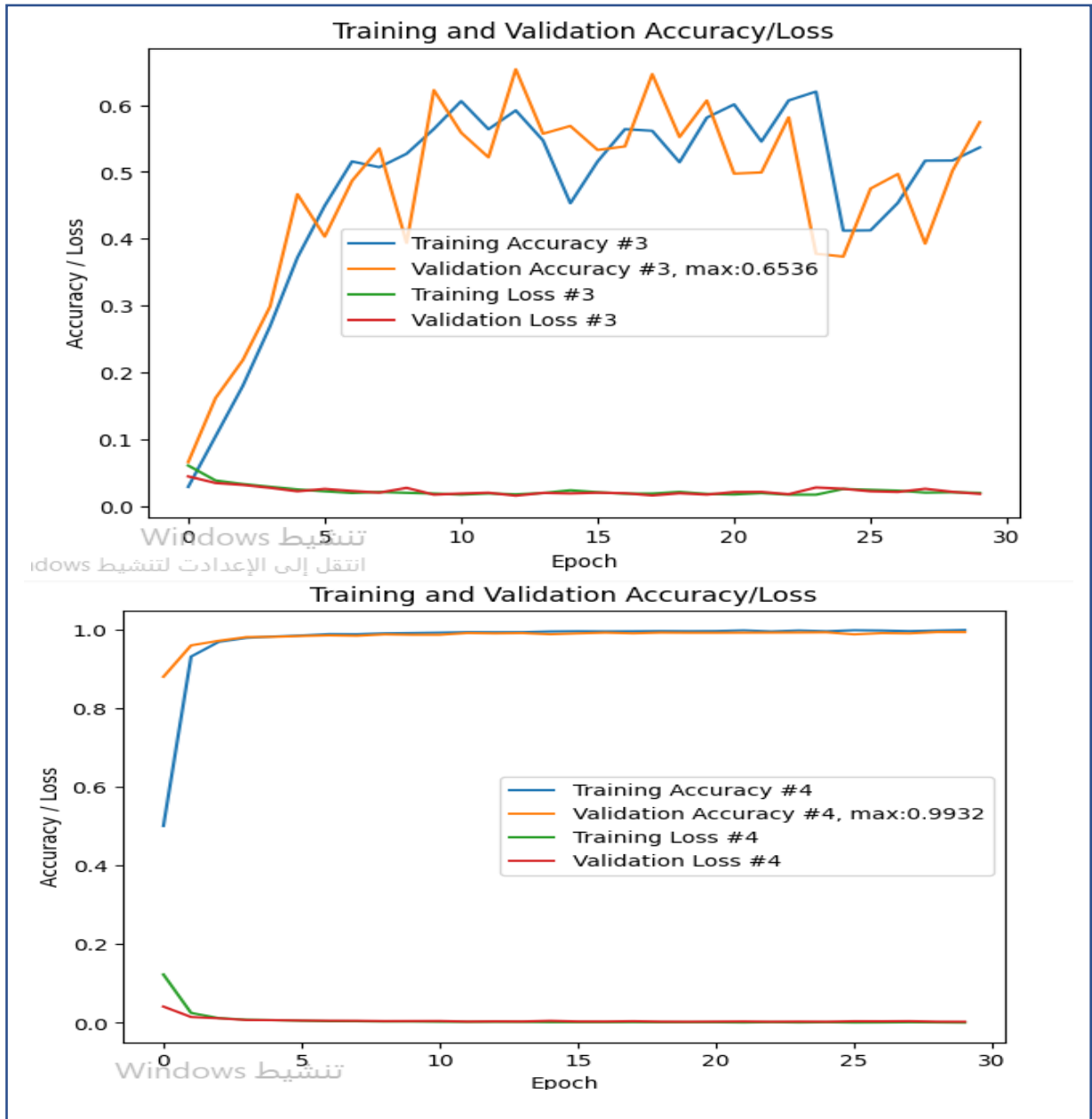
نفس النموذج على مجموعات البيانات 3,4.

- **تقليل الإفراط في التدريب** يمكن آليات البوابات في LSTM من التحكم بشكل أفضل في تدفق المعلومات، مما قد يساعد في تقليل الإفراط في التدريب في مجموعات البيانات الأصغر أو عندما يكون النموذج عرضة للضجيج [71]. على عكس SimpleRNN، بسبب قيودها في التعامل مع التبعيات طويلة الأجل فهي لا



الشكل (5-12) منحنيات الدقة والخطأ عند تطبيق بنية النموذج CNN+ simpleRNN على مجموعتي البيانات 2,1

تحتوي على آليات مشابهة للبوابات مثل LSTM ، وبالتالي تكون أكثر عرضة للإفراط في التدريب وأقل قدرة على التعميم في المهام التي تتطلب تتبع العلاقات الزمنية على مدى طويل مثل بيانات الحساسات المرتبطة زمنياً.

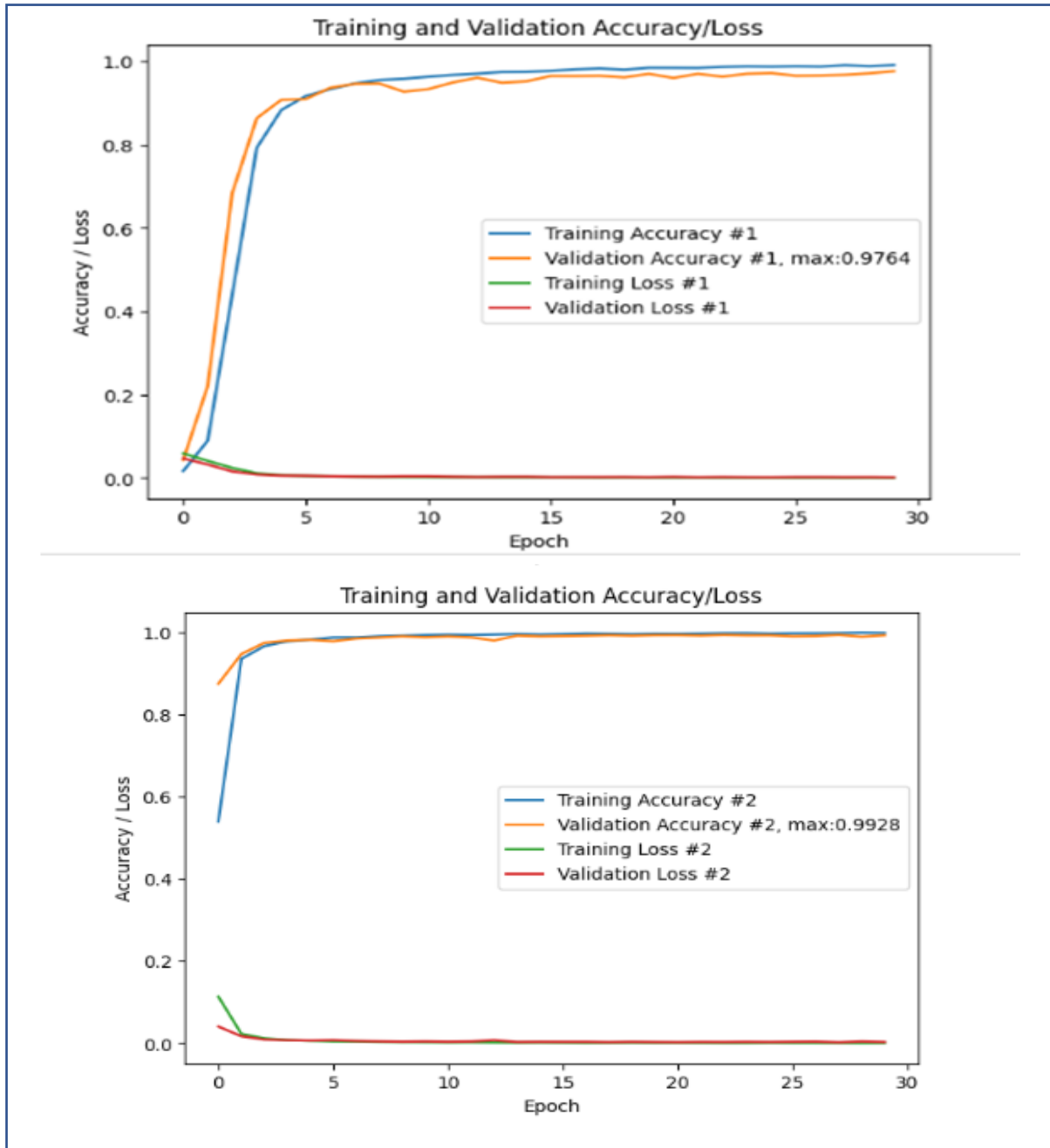


الشكل (5-13) منحنيات الدقة والخطأ عند تطبيق بنية النموذج CNN+ simpleRNN على مجموعتي البيانات 3,4

كما أنه أثناء عملية التدريب، يمكن أن تتلاشى التدرجات أو تتفجر في SimpleRNN، مما يجعل من الصعب على الشبكة تعلم التبعية طويلة المدى في البيانات.

3.3.6.5 مقارنة بين (64 vs 128) CNN+LSTM:

أدى زيادة عدد الوحدات إلى نتائج أفضل، وذلك للأسباب التالية:



الشكل (5-14) منحنيات الدقة والخطأ عند تطبيق بنية النموذج CNN+ LSTM(64) على مجموعتي البيانات 1,2

التعلم المعزز للميزات: وحدات LSTM مسؤولة عن تعلم وتمثيل الأنماط الزمنية المعقدة في البيانات المتسلسلة،

من خلال زيادة عدد الوحدات في طبقة LSTM [67]، يكتسب النموذج القدرة على النقاط التبعية الزمنية الأكثر

تعقيداً داخل بيانات الإدخال. وهذا يسمح للشبكة بالتمييز بشكل أفضل بين الأنماط المختلفة والتعرف على

الاختلافات الأكثر صعوبة في الملاحظة، مما يؤدي إلى تحسين الدقة.

- **زيادة سعة النموذج:** يؤدي العدد الأكبر من الوحدات في طبقة LSTM إلى زيادة قدرة النموذج على تعلم

وحفظ المعلومات من تسلسلات الإدخال بشكل فعال [68]. وهذا يمكّن LSTM من تخزين واستخدام المزيد

من المعلومات السابقة، وهو أمر مهم بشكل خاص للمهام التي تتطلب على تبعيات طويلة الأجل أو سياقات

زمنية واسعة النطاق. ونتيجة لذلك، يصبح النموذج مجهزاً بشكل أفضل للتعامل مع مجموعة أوسع من الأنماط

المتسلسلة المعقدة، مما يؤدي إلى دقة أعلى.

- **تجريد الميزات:** تعمل وحدات LSTM بشكل هرمي، حيث تستخرج الميزات المجردة من بيانات الإدخال أثناء

مرورها عبر الطبقات. ومع وجود عدد أكبر من الوحدات، يمكن لـ LSTM إنشاء تمثيلات ميزات قادرة على

التقاط المعلومات الأساسية والاحتفاظ بها دون ضغط أو تبسيط كبير [69]، مما يؤدي إلى دقة أفضل في إعادة

بناء البيانات التسلسلية الأصلية.

- **التعميم وتجنب الإفراط في التدريب:** على الرغم من أن زيادة عدد الوحدات تزيد من تعقيد النموذج، إلا أنها

تساهم أيضاً في تحسين أداء التعميم [70]. مع زيادة السعة، يمكن لـ LSTM تعلم تمثيلات أكثر قوة وتعميماً

أكثر تنوعاً وثراءً، مما يعزز قدرتها على التمييز بين الفئات أو الفئات المختلفة في البيانات. وهذا مفيد بشكل

خاص في المهام ذات الاختلافات الكبيرة بين الفئات، حيث يلزم وجود نموذج أكثر تعبيراً للتمييز بدقة بين

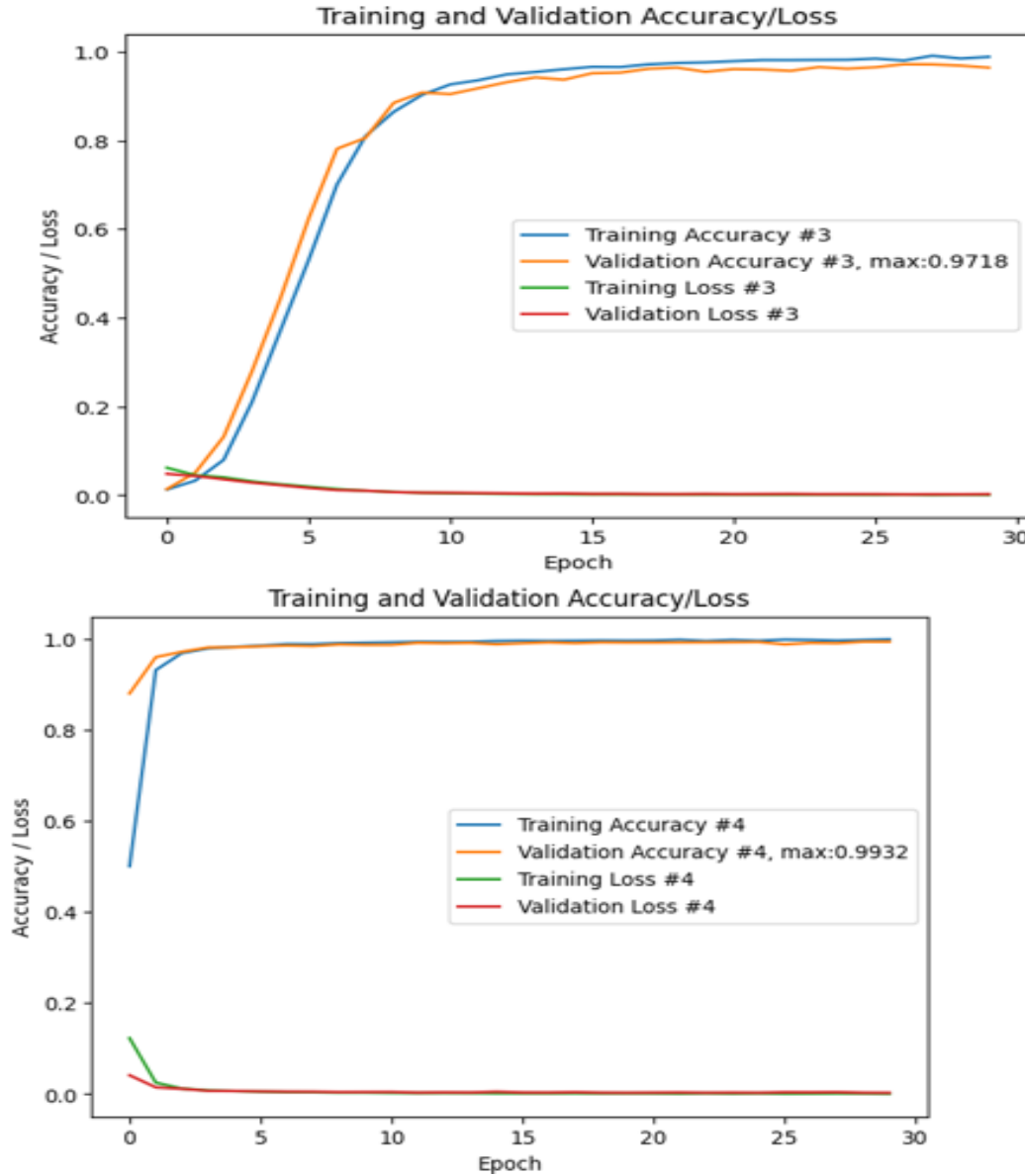
الأنماط المتشابهة.

- **تقليل فقدان المعلومات:** في النماذج التي تحتوي فيها بيانات الإدخال على تفاصيل وفوارق دقيقة، تساعد

طبقات LSTM الأكثر عدداً في تقليل فقدان المعلومات أثناء عملية التعلم [71] تمكن الوحدات الإضافية

النموذج لبيانات الإدخال، مما يساعد في تجنب الإفراط في التدريب وتحسين قدرة النموذج على الأداء الجيد

على بيانات الاختبار غير المرئية، مما يؤدي في النهاية إلى دقة أعلى في مجموعة الاختبار. يوضح الشكل (5-14) الدقة والخطأ عند تطبيق النموذج CNN+LSTM(64) على مجموعتي البيانات 2,1 والشكل (5-15) يوضح الدقة والخطأ عند تطبيق نفس النموذج على مجموعات البيانات 4,3.



الشكل (5-15) منحنيات الدقة والخطأ عند تطبيق بنية النموذج CNN+ LSTM (64) على مجموعتي البيانات 3,4

نلاحظ من المنحنيات السابقة أن المنحني الخاص بمجموعة البيانات 3 هو المنحني الذي يشير إلى الدقة

الأقل في حال تطبيق البنية CNN+SimpleRNN أو CNN+LSTM وعلى الرغم من أن مجموعتي البيانات 3,1

قد جمعنا من نفس الأشخاص الذين يبلغ عددهم 118 إلا أن الدقة عند الاختبار على مجموعة البيانات 1 كانت أفضل وذلك لأن تقسيم منحنى المشية كان باستخدام طول الخطوة في حين أنه في مجموعة البيانات 3 تم التقسيم بطول زمني ثابت وهذا يعني أن ضياع عينات المشي سوف يكون أكبر لأن الطول الثابت ربما لا يأخذ كامل سمات المشي الخاص بكل عينة عند تقسيم منحنى المشي [72].

ولكن هناك تفاوت كبير بين الدقة في حال تطبيق CNN+RNN وفي حال تطبيق CNN+LSTM على نفس مجموعة البيانات 3 حيث ارتفعت الدقة من 65.36 % إلى 97.18 % في حال استخدام CNN+LSTM كما أن المنحنى كان متعرجا بشكل كبير في حالة RNN في حين تم تنعيمه باستخدام LSTM ، يدل التعرج في كلا منحنى التدريب والاختبار على أن النموذج يواجه صعوبة في تعلم الأنماط أثناء التدريب، ولذلك أثناء الاختبار سوف يتنبأ بشكل عشوائي وبالتالي سوف يصبح المنحنى متعرج [73]. وهذا يدل على تفوق قدرات LSTM في استخراج الميزات وتعلم الأنماط على SimpleRNN.

4.6.5 مناقشة نتائج تنفيذ نموذج التعرف على الأشخاص باستخدام عينات مختلفة الطول

الدقة العالية الناتجة عن تطبيق النموذج تعزى إلى العوامل التالية:

- **بنية الشبكة العصبونية:** يجمع النموذج بين طبقة LSTM تليها طبقات تلاف 1D-Convolutional وطبقات إضافية للتقييس والتنشيط. تُعرف هذه البنية باسم الشبكة العصبونية المتعددة المقاييس LSTM Fully Convolutional Network (MLSTM-FCN)، حيث تجمع بين قوة LSTM في التعامل مع البيانات التسلسلية الزمنية [74] وقدرة الطبقات التلافية على النقاط الأنماط المحلية.
- **ضبط المعلمات الفائقة باستخدام Keras Tuner:** يتم استخدام Keras Tuner لضبط المعلمات الفائقة، وخاصة باستخدام تقنية Random Search، التي تعد تقنية فعالة لأخذ عينات من المعلمات الفائقة من مساحة محددة. يتم تحسين معلمات مثل عدد وحدات LSTM ، ومعدل الإسقاط، وحجم المرشح، وخطوة المرشح، ومعدل التعلم لتحسين أداء النموذج.

- **التنفيذ المتوازي:** إن استخدام Random Search يقوم بشكل طبيعي بعدة تجارب بالتوازي لاستكشاف مساحة المعلومات الفائقة بشكل فعال. يتيح التنفيذ المتوازي البحث السريع عن المعلومات الفائقة، وهو أمر مهم عند التعامل مع النماذج المعقدة أو مجموعات البيانات الكبيرة.
- **المكونات المخصصة:** لتعزيز القدرة التمثيلية للنموذج من خلال دمج آليات مثل إعادة تقييم الميزات Feature Recalibration. تأخذ هذه المكونات الدخل وتطبق عملية Squeeze and Excite، التي تساعد النموذج على التركيز على الميزات المهمة وتثبيط الميزات الأقل أهمية، عبر حساب متوسط التنشيط عبر القنوات باستخدام نمط التجميع Global Average Pooling. بعد ذلك، يتم إعادة تشكيل الخرج الناتج ليكون له نفس شكل الدخل.

7.5 مقارنة مع بعض الدراسات السابقة

تقدم النتائج التجريبية المقدمة في الفقرة (3.3.6.5) رؤى قيمة حول أداء وفعالية نموذج CNN+LSTM المقترح. عند الفحص الدقيق، يصبح من الواضح أن تفوق CNN+LSTM على البنات البديلة (الجدول (5-9)) يوضح المقارنة بين نتائج تطبيق نموذج CNN+LSTM (128) والدراسات السابقة) ينبع من قدرته على النقاط واستخدام الديناميكيات الزمنية المتأصلة في بيانات IMU بشكل فعال.

من خلال الاستفادة من مكون LSTM، يُظهر النموذج كفاءة ملحوظة في نمذجة الأنماط المتسلسلة، مما يسمح بفهم أكثر دقة للحركات المتطورة للأفراد بمرور الوقت. يثبت هذا السياق الزمني أنه أمر بالغ الأهمية في مهام التعرف على الأشخاص، حيث يكون تحديد الاختلافات الدقيقة في أنماط الحركة أمرًا بالغ الأهمية.

يتيح هذا التعلم الهرمي للنموذج تمييز الأنماط والعلاقات المعقدة داخل بيانات IMU، مما يؤدي إلى تحسين أداء التعرف. بالإضافة إلى ذلك، يساهم دمج تقنيات الاستنتاج الدفعي في استقرار ومتانة النموذج، مما يخفف من مشكلات مثل الإفراط في التدريب والتحول الداخلي للمتغيرات. بشكل عام، تؤكد النتائج التجريبية على

أهمية النمذجة الزمنية واستخراج السمات الهرمية في تعزيز دقة وقدرة التعميم لنماذج CNN+LSTM لمهام التعرف على الأشخاص.

الجدول (8-9) المقارنة بين العمل المطروح والدراسات السابقة

الدراسة البحثية	هذا البحث	(Hasan et al, 2022)	(Hasan et al, 2017)	(Gadaleta et al, 2018)	(Huang et al, 2021)
مجموعة البيانات 1	97.67%	95.16%	94.71%	92.91%	93.52%
مجموعة البيانات 2	99.38%	98.10%	97.67%	96.78%	97.33%
مجموعة البيانات 3	97.98%	95.22%	95.09%	-	-
مجموعة البيانات 4	99.36%	98.86%	98.58%	-	-

حيث أن التقنيات الأخرى التي جرى المقارنة معها تستخدم CNN مع تقنيات مختلفة من أجل التقاط التغيرات الزمنية والمحافظة عليها إلا أن استخدام CNN+LSTM أبدى تفوقاً في التقاط هذا التغيرات.

8.5 خاتمة الفصل الخامس : قدم هذه الفصل نماذج معززة لاستخراج المشية والتعرف عليها باستخدام

أجهزة استشعار الهواتف الذكية (IMU: وحدة القياس بالقصور الذاتي) وتقنيات التعلم العميق، أظهرت النتائج أن الاعتماد على بنية شبيهة ببنية U-Net قادرة على استخلاص ميزات قوية للتمييز بين فترات المشي والسكون بدقة جيدة ومناسبة لهذا النوع من التطبيقات كما بينت النتائج أن الدمج بين تقنيات RNN و CNN مفيدة في استخراج ميزات أفضل وبالتالي الحصول على دقة أفضل وفي النهاية يتم تقديم نموذج للتعرف على الأشخاص باستخدام عينات مشي مختلفة الطول حيث يتم تنظيم البيانات في البداية في جداول، ثم يتم تحويلها إلى صور بأبعاد (128*6)، حيث يمثل كل صف فترة حركة المشي ويتوافق كل عمود مع مكون استشعار (3 جيروسكوبات و3 مقاييس تسارع). من خلال عملية ضبط المعلمات المكثفة، يتم تحديد القيمة المثلى للمعامل 'n' (طول العينة)

لتكون 128. ثم يتم إدخال تمثيلات الصور هذه للبيانات المنظمة في نموذج التعلم العميق الذي يشتمل على طبقات التلافية ووحدات ذاكرة (LSTM). ومن ثم تمت مناقشة النتائج التي حصلنا عليها من تطبيق النماذج السابقة وأخيراً قارنا بين نموذج LSTM(128)+CNN وبعض الدراسات المنشورة.

الاستنتاجات العامة

ركزت هذه الأطروحة على تطوير وتعزيز أنظمة التعرف على الأشخاص استنادًا إلى نماذج المشية وخصائص الحركة، ومعالجة العديد من التحديات الحرجة المتعلقة باختلافات وضعية الإنسان، وتغييرات وضع الكاميرا. يمكن تلخيص المساهمات والنتائج الأساسية على النحو التالي:

شبكة عصبية التلافيفية معدلة للتعرف على الأشخاص باستخدام ميزات المشية: تم تكييف وتحسين شبكة عصبية ملتوية عميقة للتعرف على الأشخاص استنادًا إلى ميزات المشية. من خلال ضبط معلمات شبكة عصبية ملتوية عميقة وتكييف تصميم الشبكة، بما في ذلك أنواع الطبقات واستراتيجيات التطبيع، حقق النموذج تحسينات كبيرة في دقة التعرف. كما عزز استخدام زيادة الصورة (IA) من قوة النموذج، خاصة عند التعامل مع الاختلافات في جودة الصورة والدقة والظروف مثل الأفراد الذين يحملون أشياء. أظهرت التجارب على مجموعة بيانات Market أن زيادة الصورة زادت الدقة من 82% إلى 96.23%.

تم تطبيق نهج نقل التعلم باستخدام محول الرؤية (ViT) من أجل التعرف على الأشخاص. من خلال الاستفادة من قدرات استخراج السمات العامة والمحلية لمحول الرؤية (ViT)، أظهر النموذج أداءً استثنائيًا في مهام التعرف. أسفرت النتائج عن دقة بنسبة 98.43% على مجموعة بيانات Market-1501. تسلط هذه النتائج الضوء على قوة محول الرؤية (ViT) في التقاط الخصائص التفصيلية للأفراد، مما يجعله فعالاً للغاية للمهام البصرية المعقدة. أدى استخدام شبكة U-Net إلى نتائج جيدة في فصل فترات المشي عن فترات الراحة حيث وصلت الدقة إلى 86,45%.

نظام IMU-GR المحسن للتعرف على الأشخاص: استخدم نظام IMU-GR المحسن المقترح وحدات القياس بالقصور الذاتي (IMUs) من الهواتف الذكية لاستخراج أنماط المشي والتعرف عليها في بيئات مختلفة. أدى الجمع بين شبكات CNN و LSTM، جنبًا إلى جنب مع ضبط المعلمات المكثف، إلى أداء مثير للإعجاب عبر مجموعات بيانات متعددة، مع معدلات دقة تتراوح من 97.67% إلى 99.38%. تؤكد قوة النظام ودقته العالية عبر مجموعات البيانات المختلفة على إمكانياته للتطبيقات في العالم الحقيقي في سيناريوهات التعرف على الأشخاص المتنوعة.

التعرف على الأشخاص بناءً على أجهزة استشعار الهواتف الذكية: تم تطوير نهج جديد باستخدام البيانات

المنظمة من أجهزة استشعار الهواتف الذكية، مثل الجيروسكوبات وأجهزة قياس التسارع، للتعرف على

الأشخاص بناءً على المشية. من خلال تحويل بيانات المستشعر إلى صور وإدخالها في نموذج التعلم

العميق، تم تحقيق دقة عالية (98.23%). تم ضبط النموذج لتحقيق الأداء الأمثل، والاستفادة من وحدات LSTM والطبقات التلافيفية والمعالجة المتوازية لتعزيز استخراج السمات من بيانات المستشعر المحدودة.

الآفاق المستقبلية

يتيح البحث في هذه الأطروحة تقدماً كبيراً في مجال التعرف على الأشخاص باستخدام نماذج المشي وميزات الحركة، وخاصة في ظروف العالم الحقيقي الصعبة. إن الأساليب التي تم تطويرها، بما في ذلك تكيف CNN، واستخدام ViT للتعرف، والأساليب الجديدة التي تستفيد من بيانات مستشعرات الهواتف الذكية، توفر أساساً قوياً لتحسين دقة التعرف وموثوقية النظام.

يمكن أن يركز العمل المستقبلي على استكشاف المزيد من تكامل نماذج الذكاء الاصطناعي المختلفة، مثل الأساليب الهجينة التي تجمع بين CNN و ViT وتقنيات التعلم العميق الأخرى. بالإضافة إلى ذلك، فإن تطبيق هذه الأساليب على مجموعات بيانات أكبر وأكثر تنوعاً، بالإضافة إلى أنظمة التعرف في الزمن الحقيقي، يمكن أن يساعد في تحسين وتوسيع نطاق تطبيق الحلول المقترحة. كما أن استكشاف استراتيجيات زيادة البيانات الجديدة وتحسين أداء النموذج في البيئات ذات الموارد المنخفضة سيكون أيضاً طرقة قيمة لمزيد من التحقيق.

يعتبر النهج المقدم في هذا البحث للتعرف على الأشخاص باستخدام مشيتهم نهجاً تجريبياً، وبهذا يمكن تجريب العديد من الخوارزميات والتقانات ومقارنة الأداء، وبما أن مجال التعلم العميق مجالاً يتم تطويره يومياً فإنه من الممكن تجريب الخوارزميات الجديدة مثل 3D Convolutional Neural Network.

يمكن اختبار النماذج المقترحة في هذا البحث على مجموعات بيانات أخرى فهناك العديد من مجموعات البيانات التي تحتوي سمات مشي مخصصة للمرضى أو لذوي الإعاقة بالجهاز الحركي، أو تكون سمات المشي تختلف باختلاف ميل الطريق الذي يمشي عليه الشخص وإنه من المفيد تطوير النماذج المقدمة في البحث كي يتم اختبارها على مثل هذه المجموعات.

كما يمكن جمع مجموعة بيانات باستخدام كاميرا توضع أمام جامعة دمشق وتدريبها باستخدام النماذج المقترحة في هذا البحث أو جمع مجموعة بيانات خاصة بالأطفال من خلال كاميرا موضوعة في روضة وذلك لأن سمات المشي الخاصة بالأطفال تختلف عن الكبار.

وفي الختام، قدمت هذه الأطروحة مساهمات كبيرة في التعرف على الأشخاص باستخدام ميزات المشي والحركة، مع تحسينات واضحة في الدقة والمتانة عبر إعدادات مختلفة. إن الأساليب والنماذج التي تم تطويرها تبشر بالخير لتطبيقات أوسع، بما في ذلك المراقبة والأمن والتعريف البيومتري، مما يوفر تقدماً في الاستخدام العملي لأنظمة التعرف القائمة على المشي.

الأبحاث المنشورة

1- نشرت مقالة أولى بعنوان:

"Analysis and best parameters selection for person recognition based on gait model using CNN algorithm and image augmentation"

(Journal of Big of Big Data, 03 January 2021), DOI:[https://doi.org](https://doi.org/10.1186/s40537-020-00387-6)

/10.1186/s40537-020-00387-6

2- نُشرت مقالة ثانية بعنوان:

" نهج التعلم العميق لاستخراج وتمييز المشي القائم على الهاتف الذكي"

جرى تحكيمها وقبولها للنشر في مجلة جامعة دمشق للعلوم الهندسية برقم /14375/

وتاريخ 2024/09/11/

3- نشرت مقالة ثالثة بعنوان:

"تعزيز نظام التعرف على المشي باستخدام حساسات الهواتف الذكية (IMU) وتقنيات التعلم العميق." جرى

تحكيمها وقبولها للنشر في سلسلة العلوم الهندسية في المجلد (46) العدد (5) من مجلة جامعة تشرين

للبحوث والدراسات العلمية برقم بتاريخ 2024/9/1/.

4- نشرت مقالة رابعة بعنوان:

"التعلم الانتقالي للمحول البصري من أجل إعادة تحديد الهوية للأشخاص استناداً إلى ميزات المشي" جرى

تحكيمها وقبولها للنشر في مجلة جامعة دمشق للعلوم الهندسية برقم /15074/ وتاريخ 2024/12/17.

قائمة المراجع Reference List

1. Ghaleb, A., & Amara, N. (2016). Soft and hard biometrics for the authentication of remote people in front and side views. *International Journal of Applied Engineering Research*, 11(14), 8120–8127.
2. Jain, A. K., Ross, A., & Pankanti, S. (2006). Biometrics: A Tool for Information Security. *IEEE Transactions on Information Forensics and Security*, 1(2), 125–143.
3. Zhang, Y., Huang, Y., Wang, L., & Yu, S. (2019). A comprehensive study on gait biometrics using a joint CNN-based method. *Pattern Recognition*, 93(1), 228–236.
4. Strukova, O., Shiripova, L., & Myasnikov, E. (2018). Gait analysis for person recognition using principal component analysis and support vector machines. *CEUR Workshop Proceedings*, 2210, 170–177.
5. Hahnel, M., Klunder, D., & Kraiss, K.-F. (2004). Color and texture features for person
6. recognition. In 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541) (Vol. 1, pp. 647–652).
7. Kawakami, T. (2020). Coronavirus gives China more reason to employ biometric tech. *Nikkei Asian Review*. <https://asia.nikkei.com/Business/China-tech/Coronavirus-gives-China-more-reason-to-employ-biometric-tech>.
8. Castro, F. M., Marín-Jiménez, M. J., Guil, N., & De La Blanca, N. P. (2017). Automatic learning of gait signatures for people identification. In *International Work-Conference on Artificial Neural Networks*, 257–270.
9. Yousef, R., Khalil, A., Samra, A., & Ata, M. (2023). Model-based and model-free deep features fusion for high-performance human gait recognition. *The Journal of Supercomputing*, 79, 12815–12852.
10. Nixon, M. S., Tan, T., & Chellappa, R. (2006). *Human Identification Based on Gait* (1st ed.). Springer.

11. Sivarathinabala, M., & Abirami, S. (2019). AGRS: Automated gait recognition system in a smart environment. *Journal of Intelligent & Fuzzy Systems*, 36(3), 2511–2525.
12. Liao, R., Yu, S., An, W., & Huang, Y. (2020). A model-based gait recognition method with body pose and human prior knowledge. *Pattern Recognition*, 98, 107069.
13. Wan, C., Wang, L., & Phoha, V. V. (2018). A survey on gait recognition. *ACM Computing Surveys (CSUR)*, 51(5), 1–35.
14. Wu, X., An, W., Yu, S., Guo, W., & García, E. B. (2019). Spatial-temporal graph attention network for video-based gait recognition. In *Asian Conference on Pattern Recognition*, 274–286.
15. Elharrouss, O., Almaadeed, N., Al-Maadeed, S., & Bouridane, A. (2020). Gait recognition for person re-identification. *Journal of Supercomputing*. <https://doi.org/10.1007/s11227-020-03409-5>.
16. Zhang, Z., Tran, L., Yin, X., Atoum, Y., Liu, X., Wan, J., & Wang, N. (2019). Gait recognition via disentangled representation learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4710–4719.
17. Lopez-Nava, I. H., Garcia-Constantino, M., & Favela, J. (2019). Recognition of gait activities using acceleration data from a smartphone and a wearable device. In *Multidisciplinary Digital Publishing Institute Proceedings*, 31(1), 60.
18. Janiesch, C., Zschech, P., & Heinrich, K. (2021, March 19). Machine learning and deep learning. *Electronic Markets*. <https://doi.org/10.1007/s12525-021-00475-2>.
19. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.
20. Fleuret, F. (n.d.). *The little book of deep learning*. <https://fleuret.org/public/lbdl.pdf>.
21. Sun, S., Chen, W., Wang, L., Liu, X., & Liu, T. (2016). On the depth of deep neural networks: A theoretical view. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)*.

22. Deng, L., & Yu, D. (2013). Deep learning: Methods and applications. *Foundations and Trends® in Signal Processing*, 7(3), 197–387.
23. Moshayedi, A., Roy, A., Kolahdooz, A., & Shuxin, Y. (2022). Deep learning application pros and cons over algorithm. *EAI Endorsed Transactions on AI and Robotics*, 1(1), 1–13.
24. Berthold, M., & Hand, D. J. (2003). *Intelligent data analysis* (2nd ed.). Springer.
25. Li, Y., Liu, D., Li, H., Li, L., Li, Z., & Wu, F. (2018). Learning a convolutional neural network for image compact-resolution. *IEEE Transactions on Image Processing*, 28(3), 1092–1107.
26. Sagawa, R., Shiba, Y., Hirukawa, T., Ono, S., Kawasaki, H., & Furukawa, R. (2016). Automatic feature extraction using CNN for robust active one-shot scanning. In *2016 23rd International Conference on Pattern Recognition (ICPR)* (234–239).
<https://doi.org/10.1109/ICPR.2016.7899639>.
27. Haykin, S. (2014). *Neural networks and learning machines* (3rd ed.). Springer.
28. Bishop, C. M. (2006). *Pattern recognition and machine learning* (1st ed.). Springer.
29. Strang, G. (2005). *Linear algebra and its applications* (4th ed.). Cengage Learning.
30. Taye, M. M. (2023). Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions. *Computation*, 11(3), Article 52.
<https://doi.org/10.3390/computation11030052>.
31. Aggarwal, C. C. (2023). *Neural networks and deep learning* (2nd ed.). Springer.
32. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning* (1st ed.). MIT Press.
33. Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., et al. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, 77, 354–377. <https://doi.org/10.1016/j.patcog.2017.10.013>.
34. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., & Kaiser, L. (2017). Attention is all you need. In *Proceedings of the 31st Conference on Neural*

Information Processing Systems (NeurIPS 2017), Long Beach, CA, USA.

<https://doi.org/10.5555/3295222.3295343>.

35. Ni, H., Li, Y., Gao, L., Shen, H. T., & Song, J. (2023). Part-aware transformer for generalizable person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 11280–11289). IEEE.

<https://doi.org/10.1109/ICCV48922.2023.01107>.

36. Ma, H., Li, X., Yuan, X., & Zhao, C. (2023). Denseformer: A dense transformer framework for person re-identification. *IET Computer Vision*, 17(5), 527–536.

<https://doi.org/10.1049/cvi2.12172>.

37. Sarker, P. K., Zhao, Q., & Uddin, M. K. (2024). Transformer-based person re-identification: A comprehensive review. *IEEE Transactions on Intelligent Vehicles*, 9(2), 45–60.

38. Sharma, C., Kapil, S. R., & Chapman, D. (2021). Person re-identification with a locally aware transformer. *arXiv:2106.03720*. <https://arxiv.org/abs/2106.03720>.

39. Rothman, D. (2022). *Transformers for natural language processing* (2nd ed.). Packt Publishing

40. Das, S., Tariq, A., Santos, T., Kantareddy, S., & Banerjee, I. (2023). Recurrent neural networks (RNNs): Architectures, training tricks, and introduction to influential research. In *Machine learning for brain disorders* (pp. 117–123). Springer.

41. Chollet, F. (2017). *Recurrent Neural Networks*. In *Deep Learning with Python* (pp. 235–267). Manning Publications.

42. Zaccane, G., Karim, M. R., & Menshaw, A. (2018). LSTM networks. In *Deep learning with TensorFlow* (pp. 167–189). O'Reilly Media.

43. Ilemobayo, J., & Duorodola, O. (2024). Hyperparameter tuning in machine learning: A comprehensive review. *Journal of Engineering Research and Reports*, 26(6), 388–395.

44. Wu, Y., & He, K. (2018). Group normalization. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 3–19).
45. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015)* (pp. 234–241).
46. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 1–48. <https://doi.org/10.1186/s40537-019-0197-0>.
47. Zhuang, F., Qi, Z., & Duan, K. (2020, June). A comprehensive survey on transfer learning. *arXiv:1911.02685v3*. <https://arxiv.org/abs/1911.02685v3>.
48. Market Dataset. (n.d.). *Market dataset*. <https://www.aitribune.com/dataset/2018051063>.
49. Anderson, B., Shi, M., Tan, V. Y., & Wang, Y. (2019). Mobile gait analysis using foot-mounted UWB sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(3), 1–22. <https://doi.org/10.1145/3351277>.
50. Ahmad, N., Ghazilla, R. A. R., Khairi, N. M., & Kasi, V. (2013). Reviews on various inertial measurement unit (IMU) sensor applications. *International Journal of Signal Processing Systems*, 1(2), 256–262. <https://doi.org/10.12720/ijsp.1.2.256-262>
51. Bertoli, M., Cereatti, A., Trojaniello, D., Avanzino, L., Pelosin, E., Del Din, S., Rochester, L., Ginis, P., Bekkers, E. M., Mirelman, A., & Hausdorff, J. M. (2018). Estimation of spatio-temporal parameters of gait from magneto-inertial measurement units: Multicenter validation among Parkinson, mildly cognitively impaired, and healthy older adults. *Biomedical Engineering Online*, 17(1), 1–14. <https://doi.org/10.1186/s12938-018-0536-6>.
52. Zou, Q., Wang, Y., Wang, Q., Zhao, Y., & Li, Q. (2020). Deep learning-based gait recognition using smartphones in the wild. *IEEE Transactions on Information Forensics and Security*, 15, 3197–3212. <https://doi.org/10.1109/TIFS.2020.2992783>.

53. Yoon, D. H., Kim, J. H., Lee, K., Cho, J. S., Jang, S. H., & Lee, S. U. (2024). Inertial measurement unit sensor-based gait analysis in adults and older adults: A cross-sectional study. *Gait & Posture*, 107(1), 212–217.
54. Zaki, T. H. M., Sahrim, M., Jamaludin, J., Balakrishnan, S. R., Asbulah, L. H., & Hussin, F. S. (2020, February). The study of drunken abnormal human gait recognition using accelerometer and gyroscope sensors in mobile application. In *Proceedings of the 2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA)* (pp. 151–156). IEEE.
55. Iwama, H., Okumura, M., Makihara, Y., & Yagi, Y. (2012). The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition. *IEEE Transactions on Information Forensics and Security*, 7(5), 1511–1521.
56. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *Proceedings of the International Conference on Learning Representations (ICLR)*. <https://openreview.net/forum?id=YkzDqv1yz2o>.
57. Di Russo, A., Stanev, D., Armand, S., & Ijspeert, A. (2021). Sensory modulation of gait characteristics in human locomotion: A neuromusculoskeletal modeling study. *PLoS Computational Biology*, 17(5).1–33. <https://doi.org/10.1371/journal.pcbi.1008594>
58. Tran, L., & Choi, D. (2020). Data augmentation for inertial sensor-based gait deep neural network. *IEEE Access*, 8, 12364–12378. <https://doi.org/10.1109/ACCESS.2020.3003955>.
59. Hasan, M. A. M., Al Abir, F., Al Siam, M., & Shin, J. (2022). Gait recognition with wearable sensors using modified residual block-based lightweight CNN. *IEEE Access*, 10, 42577–42588. <https://doi.org/10.1109/ACCESS.2022.3187807>

60. Zou, Q., Ni, L., Wang, Q., Li, Q., & Wang, S. (2017). Robust gait recognition by integrating inertial and RGBD sensors. *IEEE Transactions on Cybernetics*, 48(4), 1136–1150.
<https://doi.org/10.1109/TCYB.2017.2726258>.
61. TensorFlow. (n.d.). *tf.keras.layers.Conv2DTranspose*. TensorFlow. https://www.tensorflow.org/api_docs/python/tf/keras/layers/Conv2DTranspose.
62. Bai, Y. (2022). ReLU–function and derived function review. *SHS Web of Conferences*, 144, 206–211. <https://doi.org/10.1051/shsconf/202214402006>
63. Alkhatib, R. (2021). Artificial Neural Network Activation Functions in Exact Analytical Form (Heaviside, ReLU, PReLU, ELU, SELU, ELiSH). *International Journal of Computer Applications*, 3(6), 212–223.
64. Sharma, S., & Sharma, S. (2020). Activation Functions in Neural Networks. *International Journal of Engineering Applied Sciences and Technology*, 4(12), 310–316.
65. Noreen, N., Palaniappan, S., Qayyum, A., Ahmad, I., Imran, M., & Shoaib, M. (2020). A deep learning model based on concatenation approach for the diagnosis of brain tumor. *IEEE Access*, 8(1), 55135–55144.
66. Hu, J., Shen, L., & Sun, G. (2018). Squeeze–and–excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7132–7141).
67. TensorFlow. (n.d.). *Keras Tuner*. https://www.tensorflow.org/tutorials/keras/keras_tuner.
68. Shi, L. F., Liu, Z. Y., Zhou, K. J., Shi, Y., & Jing, X. (2023). Novel deep learning network for gait recognition using multimodal inertial sensors. *Sensors*, 23(2), 849–865.
69. Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., & Hoi, S. C. (2021). Deep learning for person re–identification: A survey and outlook. *IEEE transactions on pattern analysis and machine intelligence*, 44(6), 2872–2893.

70. Taha, K., Yoo, P. D., Al-Hammadi, Y., Muhaidat, S., & Yeun, C. Y. (2024). Learning a deep feature clustering model for gait-based individual identification. *Computers & Security*, 136(1), 1–19.
71. Gorgolis, N., Hatzilygeroudis, I., Istenes, Z., & Gyenne, L. G. (2019, July). Hyperparameter optimization of LSTM network models through genetic algorithm. In 2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA) (pp. 1–4). IEEE
72. Burges, B., & Oraibi, Z. (2024, January 26). Gait Recognition Using Hybrid LSTM–CNN Deep Neural Networks. *Journal of Image and Graphics*, 12, 168–175.
73. Tran, L., & Hoang, T. (2021). Multi-Model Long Short-Term Memory Network for Gait Recognition Using Window-Based Data Segment. *IEEE Access*, 9(1), 160877–160889
74. Liu, K., & Ji, S. (2023). A Novel Gait Phase Recognition Method Based on DPF–LSTM–CNN Using Wearable Inertial Sensors. *Sensors*, 23(13), 1–17.

Syrian Arab Republic

Ministry of Higher Education and Scientific Research

Damascus University

Faculty of Mechanical and Electrical Engineering

Department of Electronics and Communications Engineering



Improving Person Recognition Algorithms Using Gait Models

A Dissertation submitted in Partial Fulfillment of the Requirement for the Degree
of PhD in Advanced Communications Engineering

By

Eng.Abeer Saleh

Scientific Supervisor

Dr.Talal Hammoud

Co-Supervisor

Dr.Hazem Wannous

Academic Year

2024-2025