

طريقة جديدة لكشف سقوط الإنسان باستخدام الشبكة العصبونية الالتفافية

عمار الدقاق¹، أ. د. محمد مازن المحاييري²،

أ. د. رشا كاظم مسعود³

¹ماجستير حاسبات وتحكم آلي - قسم هندسة الحواسيب والأتمتة - كلية الهندسة الميكانيكية والكهربائية - جامعة دمشق.

²أستاذ مساعد- قسم هندسة الحواسيب والأتمتة - كلية الهندسة الميكانيكية والكهربائية- جامعة دمشق.

³أستاذ مساعد- قسم الهندسة الطبية- كلية الهندسة الميكانيكية والكهربائية- جامعة دمشق.

الملخص

ساهم انتشار تقنيات الذكاء الصناعي في العقد الأخير، ورخص الكاميرات وانتشارها الواسع، في الاهتمام بتصنيف حركات وأفعال البشر خاصة تلك الأفعال التي قد يكون لها أثر سلبي في صحتهم، وقد يؤدي في أسوأ الأحوال إلى وفاتهم، وكان لحدث السقوط الحظ الأكبر من الدراسات والبحوث، وخاصة فيما يتعلق بكبار السن، فغالباً خطر الموت نتيجة السقوط يكون من نصيبهم. تم في هذا البحث طرح طريقة جديدة - أطلق عليها تسمية صورة أمامية مكدسة بعلامات زمنية (STTFI) - Stacked Time-Tagged Foreground Image - في استخلاص صورة لهيئة الشخص تحمل مزايا زمنية ومكانية تمثل الحركة وكميتها، كما تمثل علاوة على الطرق السابقة - MHI ومثيلاتها- المزايا الساكنة المرحلية التي يمر فيها الشخص أثناء السقوط، ثم تم استخدام تلك الطريقة مع شبكة عصبونية التلافية قليلة العمق باستخدام تقنية نقل التعلم وذلك لتصنيف الحدث إلى سقوط أو عدم سقوط، وقد أعطت الطريقة نتائج تفوقت على أغلب الدراسات السابقة في الدقة حيث أعطت الشبكة دقة 99.02%، وعلى دراسات أخرى في الحساسية حيث بلغت الحساسية 100%، بينما بلغت النوعية مقدار 98.73%.

الكلمات المفتاحية: كشف السقوط، كبار السن، صورة تاريخ الحركة MHI، الشبكة

العصبونية الالتفافية CNN، التعلم العميق، نقل التعلم، الإبصار الحاسوبي.

تاريخ الإيداع: 2022/6/6

تاريخ القبول: 2022/8/8



حقوق النشر: جامعة دمشق - سورية،

يحتفظ المؤلفون بحقوق النشر بموجب

الترخيص CC BY-NC-SA 04

A New Approach For Human Fall Detection By Using A Convolutional Neural Network

Ammar Aldaqqaq¹, prof. Mohamed Mazen Almahayri², Prof. Rasha Kadem Massoud³

¹Master in computer and control, department of computer and automation engineering, Faculty of Mechanical and Electrical Engineering, Damascus University, Syria

²Assistant professor, department of computer and automation engineering, Faculty of Mechanical and Electrical Engineering, Damascus University, Syria

³Assistant professor, department of Biomedical Engineering, Faculty of Mechanical and Electrical Engineering, Damascus University, Syria

Abstract

The prevalence of artificial intelligence techniques in the last decade, and the widespread of surveillance cameras, have contributed to the classification of human motions. Especially those that may have a negative impact on their health, and may, in worst cases, lead to their death. In this paper we have introduced a new method, called STTFL, that extracts a person's silhouette with spatial and temporal characteristics representing the motions and its quantities such as previous methods (MHI, BMI, ITMI). But it also represents the spatial and temporal characteristics of static poses that person passes during a fall. Then it has been used with a low-depth convolution neural network using learning transfer technique to classify the event as fall or not fall. The method gave results that outperformed most previous studies in accuracy, and some in sensitivity. the accuracy of this method is 99.02%, sensitivity 100%, and specificity 98.73%.

Keywords: Fall Detection, Elderly People, Motion History Image (MHI), Convolution Neural Network (CNN), Deep Learning, Transfer Learning, Computer Vision, Integrated Time Motion Images (ITMI).

Received: 6/6/2022

Accepted: 8/8/2023



Copyright: Damascus University- Syria, The authors retain the copyright under a CC BY- NC-SA

المقدمة

حظي تصنيف النشاط البشري في السنوات الأخيرة باهتمام العديد من الباحثين حول العالم، خاصة مع ظهور وانتشار تقنيات الذكاء الاصطناعي التي تشق طريقها بقوة في هذه الأيام، وشمل ذلك العديد من النشاطات الحياتية اليومية *Actives of Daily Living* (ADLs) التي يقوم بها الإنسان كالمشي [1]، والركض والجلوس والانحناء والاستلقاء وركوب الدراجة [2, 3, 4]... إلخ، كما شمل الاهتمام أيضاً الحركات غير الطبيعية التي يؤديها الإنسان بإرادته كحوادث العنف والتسكع والتطفل [5, 6]، أو بدون إرادته كحوادث السقوط [5, 7, 8, 9] التي كان لها النصيب الأكبر، لما لها من علاقة بالغة على صحة الأفراد فحسب تقارير منظمة الصحة العالمية [10] يأتي حدث السقوط في المرتبة الثانية كأحد أسباب الوفاة بعد حوادث الطرق، وتتركز النسبة الأكبر في الأطفال وكبار السن؛ فالأطفال يدفعهم حب الفضول واستكشاف البيئة المحيطة بهم إلى التعرض بكثرة لحوادث السقوط التي إن لم تكن مميتة ربما سببت لهم إعاقة دائمة، أما بالنسبة لكبار السن فوق عمر 60 فإن حوادث السقوط البسيطة قد تكون مميتة لهم، فقد تتسبب بكسر في الحوض أو رضوض وجروح خطيرة تتطلب إقامة في المشفى، وفي هذه الأيام مع قلة الكوادر الطبية والتمريضية وتوابع جائحة كورونا، سيكون الأمر مرهقاً مما يتطلب مساعدة تقنية حديثة ليستطيع الكادر الطبي القليل العدد أن يقوم بمساعدة كم أكبر من المرضى، كما يمكن للمساعدة التقنية أن تغني كبير السن عن وجود مرافق دائم له في المنزل، خاصة إن كان يعيش بمفرده. لذلك انتشرت في السنوات الأخيرة العديد من الدراسات لتصميم أنظمة تستطيع اكتشاف السقوط بأسرع وقت ممكن - فكلما

كانت الفترة التي يبقى فيها الشخص مدداً على الأرض بعد السقوط طويلة زادت احتمال حدوث الوفاة [11] - وإعطاء إشارة إنذار للكادر الطبي ليقوم بالمساعدة أو إجراء اتصال بالإسعاف لتقديم المساعدة لمن يعيشون بمفردهم.

وقد صنفت تلك الطرق تبعاً لنوع الحساسات المستخدمة في عملية كشف السقوط إلى؛ طرق تعتمد الحساسات المحيطية [12, 13] وطرق تعتمد الحساسات الملاصقة للجسم [8, 14-18] وطرق جمعت بين النوعين [3] وطرق تعتمد على الإبصار الحاسوبي، والتي ظهرت في السنوات الأخيرة بسبب انخفاض كلفة الكاميرات بشكل عام وانتشارها الواسع أيضاً بين الناس من خلال الأجهزة الخلفية.

ونقسم طرق الإبصار الحاسوبي من حيث استخلاص المزايا الخاصة بالسقوط من الصور إلى طرق هندسية وإحصائية [19-27]، وطرق تعتمد على تقنيات التعلم العميق [7, 9, 28-38] التي زادت في العقد الأخير، أما بالنسبة لمرحلة التصنيف فتقسم إلى طرق تعتمد على العتبات [19, 21, 22, 23, 25, 26]، وطرق تعتمد على تعليم الآلة؛ فمنهم من اتبع الطرق التقليدية في تعليم الآلة [20, 24, 27, 37]، ومنهم من اتبع التعلم العميق [7, 9, 28-38].

وعلى اعتبار أن السقوط حركة؛ أي تغيير في موضع الشخص تابع للزمن فهو ليس صورة جامدة، إنما تتابع مجموعة من الصور على شكل فيديو يميز تلك الحركة، ورغم ذلك فقد كانت هناك أبحاث تعتمد على الصورة ذاتها في تصنيف السقوط [28, 29, 30]، إلا أن تلك الأبحاث أغفلت الجانب الزمني الذي يؤثر في نتيجة التصنيف، بينما اهتمت أبحاث أخرى بالجانب الزمني فقط واستخلاص المزايا الزمنية من سرعة وتسارع واتجاه حركة وأشهرها التدفق البصري

الفديو، ثم أدخلها إلى شبكة عصبية إلتفافية VGG-16 لتقوم الشبكة باستخلاص المزايا الحركية المتعلقة بتمييز السقوط، وبهذا يكون قد درب الشبكة على صورة حركية وليست صورة جامدة لأحد الوضعيات التي يأخذها الشخص أثناء السقوط.

قام Espinosa وآخرون [33] لزيادة دقة اكتشاف السقوط بدمج صورتي التدفق البصري الناتجتين من كاميرتين مستقلتين، ثم استخدم شبكة إلتفافية قام باقتراحها لاستخلاص المزايا الخاصة بالسقوط وتصنيفه. ولكن التدفق البصري حقيقة رديء في ملاحقة حركة الجسم البشري بسبب الضجيج والظلال وتتنوع اللباس ومعدلات الحركة المتغيرة، بالإضافة إلى كمية الحسابات الكبيرة المرافقة للتدفق البصري [40]، في حالة كشف حركة سقوط الجسم البشري استنتجت Rougier [25] أنه لا يهم اتجاه وسرعة الحركة بقدر ما يهمنا معرفة كمية الحركة، وأن صور Motion History Image (MHI) أفضل من التدفق البصري في حال كشف السقوط.

قام Dobhal وآخرون [41] باستخدام طريقة معدلة على MHI أسماها Binary Motion Image (BMI) وطبقها على شبكة إلتفافية لاستنتاج المزايا وتصنيف النشاط البشري حيث أعطت دقة 98%، ثم قامت Khraief وآخرون [35] باقتراح طريقة لكشف السقوط تعتمد على استخلاص صورتين ثنائيتي البعد، أحدهما من التدفق البصري، أما الأخرى فمستخلصة من صورة BMI، بعدها أدخلت الصورتين إلى شبكتين VGG-16 مستقلتين، طورت فيما بعد العمل إلى أربعة شبكات [11]، حيث أدخلت شبكتين إلتفائيتين إضافيتين، الأولى تأخذ صورة RGB-D، أما الشبكة الأخيرة فتعتمد على الإزاحة المشتقة من صورة التدفق البصري، حيث وصلت دقة الشبكة إلى 99.72%.

optical flow (OF) [31, 33]، وأغفلت أيضا أن الشكل الذي يأخذه الشخص أثناء السقوط يؤثر في التصنيف، بينما اهتمت أبحاث أخرى بالصور المتعاقبة لتحديد حدث السقوط [32, 34, 35, 36, 38, 11].

سنركز في هذا البحث على الطرق المتبعة في كشف السقوط والتي تجمع بين الخواص الزمنية والمكانية، وسنقترح طريقة جديدة ونطبقها على الشبكة العصبونية الالتفافية¹ convolutional neural network (CNN) لتستخلص منها المزايا وتقوم بتصنيف حدث السقوط.

1- الدراسات السابقة

بعد ظهور الشبكة العصبونية الالتفافية CNN على يد Kirzhevsky وآخرين [39] عام 2012، بدأ عصر الإبصار الحاسوبي المعتمد على التعلم العميق لاستخلاص الخصائص المعقدة للصور وتصنيفها، وكان أول استخدام لها في كشف السقوط من قبل Yu وآخرين [28] عام 2017، حيث استخدم شبكة إلتفافية مؤلفة من طبقتين لكشف أحد أربع وضعيات لهيئة² الإنسان من بينها الاستلقاء. استخدم Adhikari وآخرون [29] شبكة إلتفافية لكشف ست وضعيات للإنسان من بينها الاستلقاء دون كشف السقوط، بينما استخدمت Anishchenko [30] شبكة Alexnet الالتفافية بعد تعديلها لتحديد أحد صنفين سقوط أو عدم سقوط، حيث قامت بتدريب الشبكة من خلال طريقة نقل التعلم Transfer learning على صور كاملة للأشخاص مع الخلفية.

قام Núñez-Marcos وآخرون [31] باستخلاص صورة التدفق البصري لهيئة الشخص المستتجة من

¹ الشبكة العصبونية الالتفافية (CNN، أو ConvNet) هي نوع من الشبكات العصبونية الاصطناعية، يتم تطبيقها بشكل شائع لتحليل الصور المرئية.

² الصورة الأمامية للشخص بعد حذف خلفية الصورة

فيها شعاع مزايا مدمجاً يحوي أوصافاً عمومية وأخرى تفصيلية لأربعة وضعيات مرحلية أثناء السقوط، ليكون دخلاً لشبكة LSTM، حيث أعطت الشبكة الالتقافية دقة 98.46%، بينما أعطت شبكة LSTM دقة 98.6%.

يلاحظ من الدراسات السابقة أنها انتهجت أربعة طرق من ناحية كشف السقوط بتقنيات التعلم العميق؛ الأولى استخدام صورة واحدة لكشف السقوط من خلال تحديد وضعية الاستلقاء على الأرض، وهذا لا يعني بالضرورة السقوط، أما الطريقة الثانية فهي استخدام صورة تشتمل بمحتواها على تغيرات حركة الشخص في عدة إطارات frames من الفيديو مثل التدفق البصري، أو أن تشتمل على التغيرات المكانية والزمانية مع كمية الحركة مثل MHI، أما النهج الثالث فهو يستخدم تعدد السيول multi-streams ويكون لكل سيل شبكة إلتقافية ودخل مختلف، وهذه الطريقة تستهلك موارد الذاكرة في الحاسوب وتتطلب حسابات طويلة، والدقة التي تقدمها لا تزيد كثيراً عن الطرق السابقة، أما النهج الرابع فيعتمد شبكة عميقة مؤلفة من مرحلتين المرحلة الأولى شبكة إلتقافية تكشف المزايا المكانية في الصورة، والثانية شبكة LSTM لكشف الترابط الزماني بين تلك المزايا لتقرر النتيجة النهائية، وكذلك يعتبر هذا النهج كسالفه من حيث استهلاك موارد الحاسوب، ويبقى النهج الثاني هو الأفضل، وخاصة صور MHI ولكن المشكلة في أن تلك الصور مع أنها تعبر عن تغيرات كمية الحركة وسيرها مع الزمن إلا أنها أغفلت أن الإنسان كائن هيكلي مفصلي [42] وليس مكعباً متحركاً أو كرة صماء تتحرك كل نقاطها معاً، فإثناء السقوط أو أي حركة أخرى هناك أجزاء من الجسم أو مفاصل تبقى ساكنة، واستخدام التدفق البصري أو MHI سيحذف من تلك الصور العناصر

قام Kong وآخرون [36] باقتراح طريقة تحوي ثلاث شبكات إلتقافية من نوع Resnet-18 على التفرع؛ دخل الشبكة الأولى هي صورة هيئة الشخص Silhouette Image (SI)، أما دخل الشبكة الثانية فهو صورة MHI لهيئة الشخص، أما الصورة الثالثة فهي الصورة الديناميكية Dynamic Image (DI)، وصلت دقة شبكته إلى 98.8%.

قام Lie وآخرون [34] باستخدام صورة MHI كدخل لشبكة مؤلفة من شبكة CNN تليها شبكة Long Short-Term Memory (LSTM)³ لكشف حالة السقوط حيث تعمل شبكة الالتقافية على استخلاص الصفات المكانية والزمنية من صورة MHI بينما تقوم LSTM على مراقبة التابع الزمني لتلك الصور وذلك لكشف السقوط، وصلت دقة الشبكة إلى 97.66%.

قام Sangwan [32] وآخرون باقتراح طريقة لكشف السقوط تعتمد على صور كاميرا 360 درجة - تلتقط صورة بزواوية 360 درجة، ثم استخدم شبكة عميقة مؤلفة من جزئين متتابعين؛ الأول شبكة inception-V3 حيث يمرر لها تلك الصور، وشعاع المزايا الناتج عنها يمرر إلى شبكة LSTM لتحديد التصنيف النهائي للسقوط، أعطت شبكة الجزء الأول دقة 95.3%، بينما كانت دقة شبكة LSTM هي 91.62%.

أما Aldaqaq وآخرون [38] فقد استخدم صور الفيديو المتعاقبة للشخص كدخل لشبكة عميقة مؤلفة من جزئين متتابعين، الجزء الأول هو شبكة Alexnet بعد تعديلها، حيث استخلص من الطبقتين الأخيرتين

³ الذاكرة القصيرة طويلة المدى (LSTM) هي شبكة عصبية اصطناعية متكررة (RNN) تستخدم في مجال التعلم العميق، إذ يمكنها معالجة الصور، ويمكنها أيضاً معالجة سلاسل المعطيات (مثل الكلام أو الفيديو).

القديمة ضمن الصورة MHI، حيث تضمحل قيم بكسلات الإطارات الأقدم مع مرور الزمن، ويبين الشكل (1) صورة الشخص بعد أن رفع ذراعيه وصورة MHI المرافقة حيث نلاحظ اختفاء جسمه الساكن.

إن صور MHI جيدة في تمثيل كمية الحركة وتحديد اتجاهها، ولكنها أهملت أن الجسم البشري تكون فيه أجزاء ساكنة أثناء الحركة وتلعب دوراً في وصف الوضعية والفعل الذي يقوم به الشخص، فحدث السقوط الذي نحن بصدد اقتراح نظام لتصنيفه يقسم إلى نوعين - حسبما صنفه Frassetto [26] - سقوط غير مباشر وفيه يتعثر الشخص بشي فيتحرك جسده كتلة واحدة تقريباً إلى حين الارتطام بالأرض ثم يأخذ وضعيات مختلفة يتخللها أجزاء ساكنة من جسمه تبعاً للسقوط إلى الأمام أو الخلف أو أحد الجانبين، والنوع الآخر هو السقوط المباشر وفيه يسقط مغشياً عليه وتكون حركته عامودية ثم يتوقف بفعل ارتطام ركبته بالأرض أو إليته حيث يتعرض لتوقف لبرهة خلالها تتغير طريقة السقوط إلى ما يشبه السقوط غير المباشر، وهذه الوضعيات المرحلية الساكنة تلعب دوراً في تحديد طبيعة الحركة، وقد أهملت صور MHI، كما يلعب التلاشي أو الاضمحلال للإطارات القديمة دوراً سلبياً أيضاً في ضياع معالم بداية الحركة.

الثابتة والتي تمثل أوضاع مرحلية ذات شكل معين تساهم في إعطاء خصائص مكانية [38] تفيد في عملية التصنيف النهائي، لذلك سنطرح طريقة جديدة تتلافى سيئة MHI؛ المتمثلة في عدم إظهار الأجزاء الساكنة من هيئة الشخص التي تحمل مزايا الشكل والمكان.

2- مواد وطرق البحث

2-1 صورة تاريخ الحركة Motion History Image (MHI)

في هذه الفقرة سنلقي نظرة على صورة تاريخ الحركة والمعادلة المعبرة عنها، ومثيلاتها من الطرق. لقد اقترح هذه الطريقة Davis [4] وذلك في إطار تمييز حركة أطراف الجسم البشري، وتتلخص صورة تاريخ الحركة MHI في أخذ صورة الفرق $D(x, y, t)$ بين هيئتي الشخص في إطارين متتابعين أو على بعد خطوات محددة ضمن الفيديو ثم إخضاعها للمعادلة التالية لتوليد صورة H_t :

$$H_t(x, y, t) = \begin{cases} \tau & \text{if } D(x, y, t) = 1 \\ \max(0, H_t(x, y, t-1) - 1) & \text{otherwise} \end{cases} \quad (1)$$

حيث تشير τ على عدد بين 1 و 255، فتكون H_t عبارة عن صورة فيها شدة البكسل في كل نقطة منها هي تابع زمني لمراحل الحركة في تلك النقطة، وبذلك تكون البكسلات التي تنتمي للإطارات الأحدث التي حدث فيها التغير أكثر سطوعاً من تلك للإطارات



a- arms-wave



b-arms-wave MHI

الشكل (1) -a صور شخص يلوح بذراعيه، b- صورة MHI [4]

الجسم في أغلبه ساكناً وأخذاً شكلاً معيناً، فإن تلك الأجزاء ستحتوي على مزايا بصرية مكانية وزمانية يكون لها أثر كبير في تحديد نوعية الحركة؛ لذلك يجب أن تتضمن أي صورة حركة مشتقة من عدة إطارات هذه الأجزاء موسومة باللحظة الزمنية التي حدثت فيها، فإهمال هذه المزايا سيؤثر سلباً على تصنيف الحركة وبالتالي على دقة تصنيف النظام المقترح.

لحل هذه الإشكالية اقترحنا صورة حركة جديدة تدعى صورة أمامية مكدسة بعلامات زمنية Stacked Time-Tagged Foreground Image (STTFI) وتحسب هذه الصورة وفق المعادلة التالية:

$$STTFI(x, y, t) = \begin{cases} STTFI(x, y, t - 1) + KT + 1 & \text{if } f(x, y, t) = 1 \\ STTFI(x, y, t - 1) & \text{otherwise} \end{cases} \quad (5)$$

حيث تمثل $f(x, y, t)$ أحد بكسلات صور هيئة الشخص في اللحظة t وهي إما صفر أو واحد لأن صورة الهيئة ثنائية، أما K فهو رقم الإطار المختار من الفيديو، بينما T عبارة عن جزء عشري يختار بناء على عدد الإطارات الأعظمية المشكلة للصورة STTFI، فمثلاً لو كان عدد الإطارات الأعظمي هو N إطاراً فإن T تختار بحيث $T \leq \frac{1}{N}$. وبالتالي ستكون قيمة كل عنصر من الصورة الجديدة مؤلف من جزأين جزء صحيح وجزء عشري؛ الجزء العشري KT يعبر عن الوسوم الزمنية التي يمر بها عنصر الصورة في الموضع (x, y) ، بينما الجزء الصحيح فيعبر عن القيمة التراكمية للبيكسل في الإحداثي (x, y) ، وعليه تكون الصورة STTFI عبارة عن هيئة الشخص المتعاقبة المتراكمة على شكل بكسلات قيمها تعبر عن تكاثف بكسلات الهيئة، فكلما كانت هيئة الشخص أكثر ثباتاً في المكان كانت تلك القيمة للبيكسلات المكونة لهيئة الشخص أعلى قيمة، بينما يعبر امتداد

هناك أشكال أخرى لمعادلة صورة MHI، مثل الشكل الذي اقترحه Foroughi وآخرون [43] وأطلق على الصورة تسمية Integrated Time Motion Image (ITMI) وفق المعادلة التالية:

$$ITMI_t(x, y) = ITMI_{t-\Delta t}(x, y) + t \cdot f(t) \quad (2)$$

حيث $f(t)$ يعطى بالمعادلة:

$$f(t) = I(x, y, t) - I(x, y, t - \Delta t) \quad (3)$$

حيث $I(x, y, t)$ هي شدة البيكسل عند الإحداثيين (x, y) في اللحظة t ، وهي تشبه MHI مع تلافي تلاشي الإطارات السابقة، كما طرح Dobhal وآخرون [41] معادلة مشابهة ولكن ليس على صورة الفرق بل على هيئة الشخص وهي كالتالي:

$$BMI(x, y) = \sum_{t=1}^n t^2 \cdot I(x, y, t) \quad (4)$$

حيث n عدد الإطارات المشكلة لصورة BMI، حيث تكون البيكسلات الأحدث ذات إضاءة أعلى، كما في MHI مع تلافي تلاشي الإطارات السابقة. على أية حال فالطرق السابقة جميعها متشابهة في إعطاء البيكسلات الأحدث في الحدث القيم الأعلى مع عدم الاهتمام بالمواضع الساكنة أو الساكنة مرحلياً التي يمر بها الشخص أثناء الحركة.

2-2 الطريقة المقترحة

في هذه الفقرة سنطرح طريقة جديدة، كي نقوم باستخدامها في منظومة كشف السقوط التي سنقترحها في الفقرات اللاحقة.

1-2-2 صورة أمامية مكدسة بعلامات زمنية Stacked Time-Tagged Foreground Image (STTFI)

بما أن الجسم البشري أثناء الحركة الطبيعية أو غير الطبيعية كالسقوط تكون فيه أجزاء ساكنة كمفاصل الأطراف التي تعتبر نقاط ارتكاز، أو مثلاً أثناء السقوط يمر بمراحل عابرة لحظية يكون فيها

الهيئة على مساحة واسعة من الصورة عن الحركة لتلك الهيئة والقيم المتدنية للبكسلات تعبر عن مدى سرعة تلك الحركة.

يمكن توضيح ذلك من خلال تكوين صورة STTFI لوضعية السقوط لشخص من مجموعة معطيات التدريب التي قمنا بتصويرها كما في الشكل

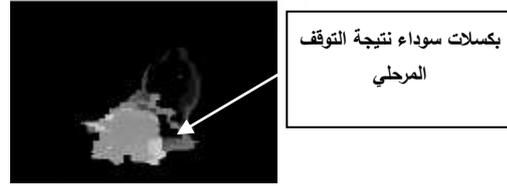
(2)



a



b



c

الشكل (2) a- مجموعة إطارات لحركة السقوط في حال فقدان الوعي، b- صورة STTFI المستنتجة منها، c- صورة MHI.

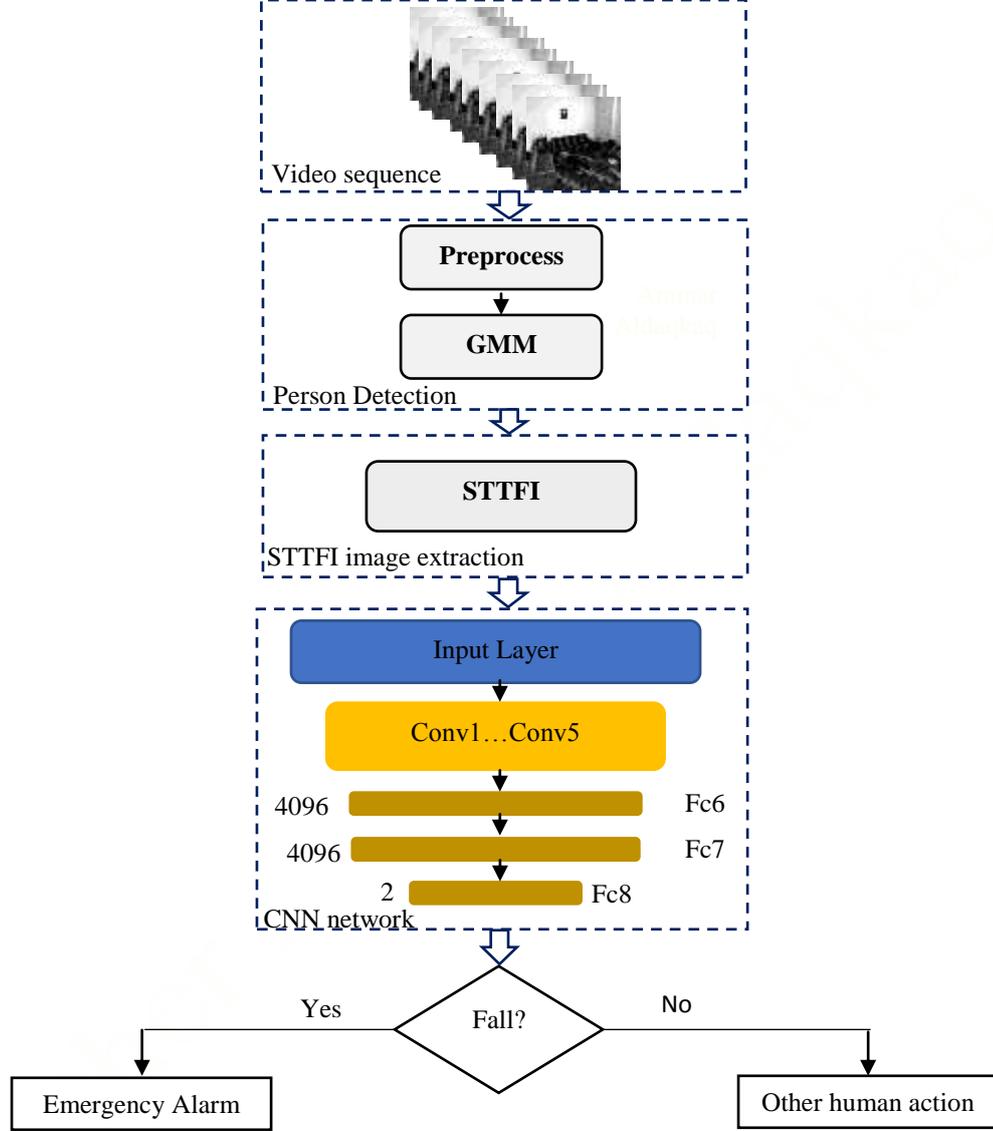
بينما نلاحظ في الصورة الجديدة STTFI الشكل (b-2) بأن الوضعيات المرحلية الساكنة واضحة تماماً، حتى في حال استقرار جسم الشخص مستقياً على الأرض نلاحظ أن بكسلات وضعية الاستلقاء الأخيرة تماثل تماماً وضعية الاستلقاء الأخيرة في الشكل (a-2) الصورة الأخيرة على اليسار، كما أنه في حالة الحركة تكون قيمة البكسلات متناسبة عكساً مع الحركة؛ فكلما كانت الحركة سريعة كانت قيمة البكسل أقل قيمة، وعلى عكس ذلك فالوضعيات شبه الساكنة أو الساكنة تكون قيمة البكسل كبيرة جداً. أما في MHI فقيمة البكسل المتحرك تعبر عن مدى حدائه فالحديث ذو شدة عالية والقديم ذو شدة منخفضة تنخفض مع قدوم إطارات جديدة لتشكيل صورة MHI.

يمثل الشكل (a-2) مجموعة من الإطارات مأخوذة من مشهد لسقوط شخص مغماً عليه من مجموعة معطيات التدريب المصورة من قبلنا FD22، أما الشكل (b-2) فيبين صورة STTFI، بينما الشكل (c-2) فيبين صورة MHI، حيث يلاحظ فيها أن الوضعيات التي يكون فيها الشخص في حالة توقف مرحلي عند اصطدام ركبتيه بالأرض وحين اصطدام إيلته بالأرض، تكون البكسلات المقابلة لها سوداء وذلك لعدم وجود حركة، ونجد ذلك أيضاً في حالة الوقوف قبل السقوط حيث يرى فقط كفاف محيط به بينما جسمه الساكن تكون بكسلاته سوداء أيضاً، وهذا يؤدي إلى فقدان المزايا المكانية المتعلقة بالشكل والتي تلعب دوراً في تحديد طبيعة الحركة.

سنستخدم صورة STTFI المقترحة الجديدة في

2-2-2 منظومة كشف السقوط المقترحة

منظومة لكشف السقوط يبينها الشكل (3)



الشكل (3) منظومة كشف السقوط المقترحة باستخدام صورة STTFI

تجميع 120 إطاراً لإرسالها للمرحلة اللاحقة، فقد وجد [38, 32] نتيجة متابعة العديد من الفيديوهات لحالات السقوط بأن أطول فترة لحدوث السقوط هي أربعة ثوان (أي قد يتعثر الشخص ثم يحاول التوازن ثم يسقط) بينما يتم السقوط الفعلي من الزلل إلى الاستلقاء في فترة تتراوح بين 300ms و 500ms، وعليه نجد أن 15 إطاراً كحد أقصى تكافئ فترة السقوط الفعلي، بينما

تتألف المنظومة من أربعة مراحل:

1-2-2-2 استحصال الأطر من الفيديو

يتم في هذه المرحلة استحصال الإطارات من الفيديو بزمن اعتيان ثابت مقداره 0.04s وهو يقابل معدل الإطار 25f/s، وبهذه الطريقة مهما كبر معدل الأطر الذي التقط فيه الفيديو، فإن الأطر المستحصلة تظهر وكأنها سجلت بمعدل إطار 25f/s، يتم بعدها

استخلاص المزايا المكانية والزمانية من الصورة، ثم بعدها تقوم بمهمة تصنيفها إلى حدث سقوط أو عدم سقوط.

وقد اعتمدنا في هذه المنظومة على شبكة Alexnet التي تتمتع ببساطة التركيب قلة العمق وقوة التمييز وهي أول شبكة اقترحت في مجال التعلم العميق من قبل Alex و Hinton [39] وأثبتت كفاءتها في تمييز الأصناف الألف التي دربت عليها.

حيث يبين الشكل (4) هذه الشبكة والتي تتألف من 5 طبقات يدعى كل منها بطبقة الطي conv أو الالتفاف الرياضي، تليها ثلاث طبقات كاملة الاتصال؛ تحتوي طبقة الطي الواحدة على مجموعة من المرشحات بأبعاد محددة، ويستخدم كل منها في عملية طي مع جزء مماثل له في الأبعاد من صورة الدخل حيث ينتج عنها خرج أحد العصبونات في خريطة المزايا المستخلصة وتكرر العملية على كل جزء من صورة الدخل حتى الحصول على خريطة المزايا (Feature map) أو الخصائص المرتبطة بهذا المرشح، تؤمن عملية الطي بهذه الطريقة التعرف على ميزات الصورة بشكل أفضل نظراً لارتباط خرج العصبون بالنقاط المتجاورة في ذلك الجزء من صورة الدخل، وهذا ما يسمح أيضاً بتمثيل أفضل للمناطق المتداخلة، طبعاً خرج طبقة الطي أو الالتفاف لا تمرر إلى الطبقة التي تليها كما هي إنما يطبق عليها تابع تنشيط رياضي (Relu) لتلافي التفاوت في قيم الخرج، يليه في بعض الطبقات تابع الانتخاب (Maxpool) والذي يقلص حجم الخرج، ويخلص الشبكة من سلبية الانطباق الزائد (Overfitting).

السقوط بشكل عام قد يستغرق كحد أقصى زمن 120 إطاراً.

2-2-2-2 كشف هيئة الإنسان Person detection

تتألف هذه المرحلة من قسمين؛ القسم الأول هو المعالجة الأولية، ويتم فيها تعديل حجم الأطر المستحصلة إلى قياس $227 \times 227 \times 3$ ، باستخدام الهرم الغوسي (Gaussian pyramid). كما يتم في هذه المرحلة تحويل الصور إلى الرمادي.

أما القسم الثاني فيتم فيه استخدام نموذج المزيج الغوسي (Gaussian Mixture Model (GMM لفصل هيئة الشخص عن خلفية الصورة. حيث نحصل في نهاية هذه المرحلة على مجموعة صور لهيئة الشخص بدون الخلفية وبحجم مناسب لتكون دخلاً للمرحلة التالية.

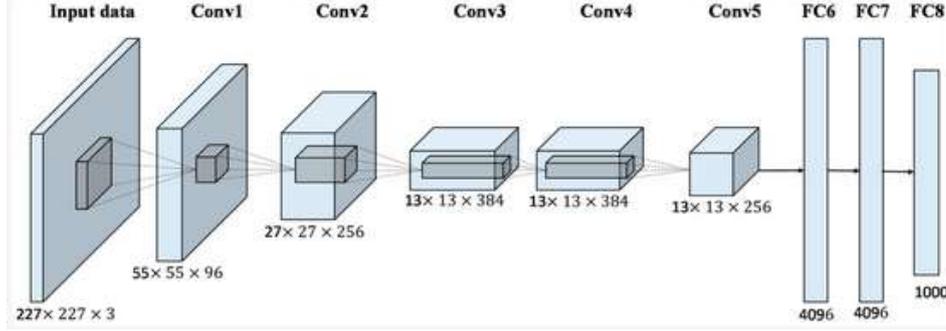
3-2-2-2 استخلاص صورة STTFI

في هذه المرحلة يتم أخذ مجموعة الصور الناتجة من المرحلة السابقة، وتشكيل صورة STTFI الجديدة المقترحة التي شرحناها سابقاً.

حيث تحتوي هذه الصورة على مزايا مكانية وزمانية ليس فقط بما يتعلق بالحركة في المشهد بل حتى بما يتعلق بالنقاط الساكنة والساكنة مرحلياً في المشهد، وبالتالي فإن هذه الصورة تحتوي مواصفات تمييزية في وصف الفعل الحركي الذي يقوم به الشخص، وهذا ما سيظهر في النتائج اللاحقة.

4-2-2-2 الشبكة العصبونية الالتفافية CNN

في هذه المرحلة يتم ادخال صورة STTFI على الشبكة العصبونية الالتفافية لكي تقوم بمهمتين الأولى



الشكل (4) بنية شبكة Alexnet الالتفافية [39]

وغيرها، فهناك مشاهد لغرفة الجلوس والمكتب ...
إلخ.

- مجموعة URFD وتتألف من 30 مشهد سقوط و 40 مشهداً للنشاطات اليومية، بأبعاد 640x480.
- مجموعة U. Montreal تحوي 24 مشهداً مسجلة بثماني كاميرات (أي كل مشهد مصور ثماني مرات كل مرة من زاوية مختلفة)، 22 مشهداً منها تتضمن السقوط، بينما مشهدها يتضمنان النشاطات الحياتية اليومية، وسجلت تلك المشاهد بمعدل إطار 120f/s وبأبعاد 720x480، وجميع تلك المشاهد مسجلة في المختبر.

- مجموعة التدريب المصورة من قبلنا FD22 وتتألف من 225 مشهداً - 47 مشهد سقوط والباقي نشاط حياتي يومي - مصورة بأبعاد 960x720 وبمعدل إطار 30f/s، وذلك في كل من غرفة الجلوس وغرفة النوم، ومن خلال متطوعين، وتحوي على 25485 صورة، منها 5297 صورة سقوط بمختلف الوضعيات، بما فيها صور جديدة للسقوط نتيجة فقدان الوعي، وكذلك السقوط من على الأريكة، ليست موجودة في المجموعات السابقة.

يبين الشكل (5) عدة صورة من تلك المجموعات

وقد قمنا بتعديل بنية الشبكة لتلائم الغرض من المنظومة وهو التعرف على حدث السقوط أو عدم السقوط، ولذلك فقد حذفنا طبقة FC8 التي تتعرف على 1000 صنف واستبدلناها بطبقة كاملة الاتصال FC8 جديدة بعصبونين للتعرف على أحد الحدثين. وعند تعليم الشبكة سيتم فقط التعليم من الصفر لهذه الطبقة بينما يتم تحديث قيم الأوزان في الطبقات الأخرى بما يتناسب مع مجموعات التدريب الجديدة وهذا جوهر نقل التعلم (transfer learning) الذي من خلاله تحافظ الشبكة على ما تعلمته سابقاً وتتعلم بأمثلة قليلة وبفترة قصيرة تصانيف جديدة.

3-2 مراحل تحضير المعطيات والتدريب

1-3-2 معطيات التدريب

لقد قمنا بتدريب الشبكة على عدة مجموعات صور تدريب وهي:

- مجموعة FDD أو ما تعرف أيضاً Le2i، تتألف هذه المجموعة من 221 مشهداً، تتضمن مختلف النشاطات اليومية بما فيها السقوط، ومسجلة بمعدل إطار 25f/s وبأبعاد 320x240، والمشاهد السابقة غنية بالخلفيات المتنوعة من مفروشات وكراسي



صورة من مجموعة الصور U. Montreal



صورة من مجموعة الصور FDD



صورة من مجموعة الصور URFD



صورتين من مجموعة الصور FD22

الشكل (5) عدة صور من مجموعات التدريب الأربعة المستخدمة جميعها في تدريب الشبكة الالتفافية

دورات التدريب (Epochs) فقد اخترناها تساوي 45 دورة، أما حجم الدفعة (Batch size) من الصور فهو 64. وقد تمت مرحلة التدريب على جزأين؛ الجزء الأول قمنا فيه بتدريب الشبكة المعدلة في المرة الأولى على صور STTFI مستنتجة من مجموعة المعطيات المصورة من قبلنا FD22، ثم أعدنا التدريب على صور MHI مستنتجة من ذات مجموعة التدريب، ثم قارنا بين نتائج الاختبار لمعرفة هل الطريقة المقترحة أفضل من طريقة MHI.

أما الجزء الثاني فقد درنا فيه الشبكة المعدلة على صور STTFI المستنتجة من مجموعات التدريب الأربعة جميعها.

3- النتائج والمناقشة

تم تقييم أداء الشبكة من خلال ثلاثة معايير أداء رئيسية وهي الدقة (Accuracy) والحساسية (Sensitivity) والنوعية (Specificity) وتعطى بالعلاقات التالية:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

2-3-2 تحضير المعطيات

في هذه المرحلة قمنا بتقسيم المشاهد إلى لقطات، ثم طبقنا المرحلتين الثانية والثالثة من الخوارزمية لنحصل في نهايتهما على صور STTFI، حيث حفظنا الصور التي تمثل أحداث السقوط في مجلد يحمل الاسم fall، بينما حفظنا الصور التي تمثل النشاط الحياتي اليومي في مجلد يحمل الاسم no_fall.

2-3-3 مرحلة التدريب

في هذه المرحلة قمنا بحذف الطبقة FC8 التي تميز بين 1000 صنف، وأضفنا طبقة FC8 مؤلفة من صنفين فقط وهما الصنفان اللذين ستميزهما شبكة Alexnet العصبونية الالتفافية، ثم قمنا بتدريب الشبكة المعدلة بتقنية نقل التعلم، باستخدام الماتلاب 2020a، وقد قسمنا مجموعة معطيات التدريب إلى 70% للتدريب و 15% للاختبار و 15% للصلاحيية (Validation)، أما بالنسبة لمعدل التعليم (Learning rate) فقد جعلناه بمقدار 0.001، أما بالنسبة لعدد

ومن الجدولين السابقين يمكن حساب كل من المعايير الثلاثة السالفة الذكر، حيث يبين ذلك الجدول (3).

الطريقة	الحساسية	النوعية	الدقة
Accuracy	Specificity	Sensitivity	
MHI	90%	92.31%	91.66%
STTFI	100%	96.15%	97.22%

نلاحظ من الجدول بأن دقة الشبكة باستخدام الطريقة المقترحة أعلى منها باستخدام طريقة MHI، حيث وصلت دقة الشبكة بالطريقة المقترحة إلى 97.22%، بينما باستخدام MHI وصلت دقة الشبكة ذاتها إلى 91.66%، وكذلك نجد أن الشبكة العصبونية الالتفافية باستخدام الطريقة المقترحة STTFI استطاعت تمييز كل حركات السقوط المختبرة، بينما الشبكة العصبونية الالتفافية باستخدام صورة MHI وصلت الحساسية فيها إلى 90% فقط، وعليه فطريقة STTFI المقترحة الجديدة أفضل في تمثيل الأفعال والحركات البشرية من طريقة MHI، وهذا ما يثبت صحة طرحنا لهذه الطريقة فهي تستطيع تمثيل الحركة التي يمر بها الجسم البشري أثناء قيامه بفعل معين، إضافة إلى تمثيل اللحظات الزمنية التي تكون فيها أجزاء من الجسم ساكنة مرحلياً أو كلياً، بينما MHI والطرق الأخرى المعدلة عنها تمثل بشكل جيد كمية الحركة واتجاهها ولا تهتم بالأجزاء الساكنة أو الساكنة مرحلياً، والتي تحمل مزايا زمنية ومكانية هامة في تحديد نوع الحركة أو الفعل البشري.

أما فيما يتعلق بالجزء الثاني من التدريب فإن الجدول (4) يبين مصفوفة الالتباس للشبكة المعدلة المدربة على المجموعات التدريبية الأربعة جميعها.

الوضع الحقيقي	عدد صور الاختبار	الوضع المُنتَبأ به	
		سقوط	عدم سقوط
السقوط	69	69	0
عدم السقوط	237	3	234

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (7)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (8)$$

حيث TP (True Positive) تعني عدد مرات السقوط المكتشفة بشكل صحيح، و TN (True Negative) عدد الصور المعبرة عن النشاط الحياتي اليومي المكتشفة بشكل صحيح على أنها عدم سقوط، أما FP (False Positive) هو عدد الصور المصنفة كسقوط بينما هي حقيقة نشاط حياتي يومي، وأخيراً FN (False Negative) وهو عدد الصور التي صنفت على أنها ليست سقوط بينما هي في الحقيقة سقوط.

وتقيس الدقة قدرة الطريقة المقترحة على التمييز بين أحداث السقوط وأنشطة الحياة اليومية، بينما تحدد الحساسية قدرة الطريقة المقترحة على تصنيف أحداث السقوط بشكل صحيح، أما النوعية فهي تحدد قدرة تلك الطريقة المقترحة على التمييز بين أنشطة الحياة اليومية بشكل صحيح.

يبين الجدول (1) مصفوفة الالتباس (confusion Matrix) المتعلقة بتعليم الشبكة الالتفافية المعدلة على صور MHI للجزء الأول من التدريب، كما يبين الجدول (2) مصفوفة الالتباس المتعلقة بتعليم الشبكة الالتفافية المعدلة على صور STTFI للطريقة المقترحة.

الجدول (1) مصفوفة الالتباس المتعلقة بشبكة Alexnet المدربة على صور MHI

الوضع الحقيقي	عدد صور الاختبار	الوضع المُنتَبأ به	
		سقوط	عدم سقوط
السقوط	10	9	1
عدم السقوط	26	2	24

الجدول (2) مصفوفة الالتباس المتعلقة بشبكة Alexnet المدربة على صور STTFI

الوضع الحقيقي	عدد صور الاختبار	الوضع المُنتَبأ به	
		سقوط	عدم سقوط
السقوط	10	10	0
عدم السقوط	26	1	25

ونلاحظ من الجدول (4) أن الدقة النهائية التي حصلنا عليها هي 99.02% والحساسية والنوعية هما 100% و 98.73% على التوالي.

ويبين الجدول (5) مقارنة الطريقة المقترحة مع الطرق المعتمدة على الشبكات الالتفافية وصور ملونة أو صور تدفق بصري.

ونلاحظ من الجدول (5) بأن الطريقة المقترحة تتفوق على جميع تلك الطرق السابقة، فرغم أن Yu وآخرون [28] وصل إلى دقة عالية 99.01% وطريقتنا أعطت دقة 99.02% أعلى بقليل من تلك الطريقة إلا أن الحساسية أي تمييز حدث السقوط بالنسبة لطريقتنا هي 100%.

وكما أنها تفوقت من حيث الدقة والحساسية على طريقة Adhikari وآخرون [29] التي استخدم فيها صور RGB-D ويعود السبب في هذا إلى قلة العينات المستخدمة إضافة إلى محدودية المزايا في صور RGB المستخدمة ونوعية شبكته الالتفافية، وأما Anishchenko [30] فقد كانت الحساسية في طريقتها منخفضة جداً لأنها استخدمت مجموعة واحدة وصور RGB فقط. أما بالنسبة Núñez-Marcos وآخرون [31] فإن طريقتنا تفوقت عليه في الدقة والحساسية والنوعية رغم استخدامه شبكة التفافية أكثر عمقا وتعقيداً. وقد تفوقت طريقتنا على طريقة Espinosa وآخرون [33] والتي استخدم فيها صور التدفق البصري التي تعبر فقط عن سرعة الحركة فقط.

الجدول (5) مقارنة بين الطريقة المقترحة والطرق التي تستخدم الشبكة العصبونية الالتفافية مع صور ملونة أو تدفق بصري

الطريقة	الشبكة المستخدمة	نوع الصور المستخدمة	مجموعة معطيات التدريب الصورية	الحساسية % Sensitivity	النوعية % Specificity	الدقة % Accuracy
Yu وآخرون (2017) [28]	شبكة CNN مقترحة	صور ثنائية اللون	مجموعة صورية خاصة به	98.98	99.03	99.01
Adhikari وآخرون (2017) [29]	شبكة CNN مقترحة	RGB-D	مجموعة صورية خاصة به	99.73	64.44	74
Anishchenko [30](2018)	Alexnet	RGB	FDD	61	99	99
Núñez-Marcos وآخرون (2017) [31]	VGG-16	OF	URFD Montreal FDD	100 99 99	92 96 97	95 - 97
Espinosa وآخرون [33](2019)	شبكة CNN مقترحة	2- OF	UP-Fall Detection	97.95	83.08	95.64
طريقتنا	Alexnet	STTFI	URFD, FDD, UMontreal, FD22 خاصة بنا	100	98.73	99.02

ويبين الجدول (6) مقارنة بين الطريقة المقترحة والطرق المعتمدة على تعدد السيول multi-stream التي يقصد بها عدة شبكات إلتفافية تعمل على التفرع معاً وتأخذ صور مختلفة الصيغ، وهذا يعني شبكة ذات حجم كبير، حيث يلاحظ أيضاً من الجدول (6) ورغم أن أضخم شبكة عصبونية التفافية والمؤلفة من أربعة سيول كما هو ظاهر مع Khraief وآخرون [11]

وصلت إلى دقة 99.72% مع مجموعة المعطيات FDD وحساسية 100% مع مجموعة المعطيات URFD، إلا أن طريقتنا المقترحة حصلت على ذات الحساسية وبشبكة Alexnet أقل عمقاً وأبسط مع مجموعة معطيات متنوعة وأضخم، كما أن النوعية في طريقتنا أعلى منها في تلك الطريقة، فقد وصلت النوعية في طريقتنا إلى 98.73%، بينما في تلك

طريقة Aldaqlaq وأخرون [38] نجد بأن الطريقة المقترحة أفضل من ناحية الدقة والنوعية، بينما كلاهما يتمتعان بتمييز كامل لحدث السقوط، حيث كلاهما له ذات الحساسية 100%، فالطريقة المقترحة تحمل ذات بنية الشبكة الالتفافية Alexnet وتستخدم فقط صور STTFI التي تحمل خصائص زمنية تتابعية ومكانية تعبر عن الشكل أثناء الحركة وأثناء لحظات التوقف المؤقت، بينما الطريقة الأخرى تستخدم شبكة LSTM لكشف التتابع لوضعيات السقوط المرحلية، وهي أعمق، وبالتالي تستهلك موارد الحاسوب بشكل أكبر. فباستخدام صور STTFI وبشبكة قليلة العمق حصلنا على دقة وحساسية أفضل من استخدام شبكات عميقة على مرحلتين.

الطريقة وصلت إلى 95% مع مجموعة المعطيات URFD، وهذا يدل أن صور STTFI تحمل صفات مكانية وزمانية أكثر مما تمثله من صفات الصور الأربعة المستخدمة في تلك الطريقة. ويمكن النظر إلى بقية الطرق في الجدول (6) ليلاحظ بأن طريقتنا المقترحة هي الأفضل بينهم، ويعود ذلك لقدرة صور STTFI على تمثيل القعل الحركي والمراحل الساكنة منه زمانياً ومكانياً.

كما يبين الجدول (7) أيضاً مقارنة بين الطريقة المقترحة والطرق التي تعتمد على شبكة عميقة مؤلفة من مرحلتين متتابعتين الأولى إلتفافية والثانية LSTM، حيث نلاحظ من الجدول بأن الطريقة المقترحة أفضل من كلا الطريقتين الموجودتين في الجدول إن كان من ناحية الدقة أو من ناحية الحساسية، فبالمقارنة مع

الجدول (6) مقارنة بين الطريقة المقترحة والطرق التي تستخدم الشبكة العصبونية الالتفافية مع تعدد السيول

الطريقة	الشبكة المستخدمة	نوع الصور المستخدمة	مجموعة معطيات التدريب الصورية	الحساسية Sensitivity %	النوعية Specificity %	الدقة Accuracy %
Khraief وآخرون [35](2019)	2-streams VGG-16	OF BMI	URFD	100	92.5	-
Kong وآخرون [36] (2019)	3-streams Resnet18	SI DI MHI	Dataset-1 خاصة به Dataset-2 خاصة به	98.8 100	98 99.3	- -
Khraief وآخرون [11] (2020)	4-streams VGG-16	OF BMI Disp.OF RGB-D	FDD URFD	- 100	- 95	99.72 -
طريقتنا	Alexnet	STTFI	URFD, FDD, UMontreal, FD22 خاصة بنا	100	98.73	99.02

الجدول (7) مقارنة بين الطريقة المقترحة والطرق التي تستخدم شبكة عصبونية عميقة مؤلفة من مرحلتين متتابعتين

الطريقة	الشبكة المستخدمة	نوع الصور المستخدمة	مجموعة معطيات التدريب الصورية	الحساسية Sensitivity %	النوعية Specificity %	الدقة Accuracy %
Sangwan وآخرون [32] (2018)	inceptionV3 and LSTM	RGB from ceiling	360°dataset خاصة به	-	-	91.62
Aldaqlaq وآخرون [38] (2021)	Alexnet and LSTM	RGB	URFD, FDD, UMontreal	100	97.22	98.6
طريقتنا	Alexnet	STTFI	URFD, FDD, UMontreal, FD22 خاصة بنا	100	98.73	99.02

تقنية التعلم العميق بواسطة الشبكات العصبونية الالتفافية، حيث اخترنا شبكة Alexnet لبساطتها وقلة عمقها ودقتها العالية في التصنيف، وقمنا بتعديلها لتستطيع التمييز بين صنفين هما السقوط وعدم السقوط، وقمنا بتصوير مجموعة معطيات تدريب خاصة FD22 تحوي وضعيات جديدة مثل حالة الإغماء، وقمنا بعدها بتدريب الشبكة بواسطة نقل التعلم على تلك الصور المستنتجة من مجموعة المعطيات الأربعة جميعها (URFD, FDD, Umontreal, FD22)، وقد أعطت نتائج أفضل من الأبحاث السابقة في هذا المجال حيث أعطت دقة 99.02% وحساسية 100% ونوعية 98.73%.

هذا البحث ممول من جامعة دمشق وفق رقم التمويل (501100020595).

وهذا ما نراه أيضا مع Sangwan [32] فدقة الطريقة المقترحة أفضل من طريقته رغم استخدامه لصور سلفية بحيث تخلص من سيئة اختفاء أجزاء من الجسم وراء الأثاث، واستخدامه شبكة عميقة inceptionV3 تلتها شبكة LSTM، إلا أن مجموعة صور كانت من نوع RGB وقليلة وتعتمد على شخص واحد.

4- الخلاصة

في هذه البحث طرحنا طريقة جديدة لكشف سقوط الإنسان بواسطة الكاميرا، تتلخص الطريقة في استخلاص صورة حركة جديدة أسميناها STTFI تختصر فيها مجموعة من إطارات الفيديو المتلاحقة، حيث تتضمن تلك الصورة معلومات زمانية ومكانية عن حركة الجسم البشري، بالإضافة إلى تمثيل الأجزاء الساكنة أو الساكنة مرحليا، بحيث تمثل كل الوضعيات المرحلية التي يمر بها الجسم البشري أثناء قيامه بحركة معينة إرادية أو لا إرادية، ثم قمنا باستخدام

المراجع References

1. Borges, P. (2013). **Pedestrian Detection Based on Blob Motion Statistics**. *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 224 - 235.
2. ULLAH, A., AHMAD, J., MUHAMMAD, K., SAJJAD, M., & W. BAIK1, S. (2017). **Action Recognition in Video Sequences using Deep Bi-Directional LSTM With CNN Features**. *IEEEAccess*, pp. 1155 - 1166.
3. Bharti, P., De, D., Chellappan, S., & Das, S. (2019, April 1). **HuMAn: Complex Activity Recognition with Multi-modal Multi-positional Body Sensing**. *IEEE Transactions on Mobile Computing*, pp. 857 - 870.
4. Bobick, A., & Davis, J. (2001, Mar). **The Recognition of Human Movement Using Temporal Templates**. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 257-267.
5. Kim, D., Kim, H., Mok, Y., & Paik, J. (2021, June 27). **Real-Time Surveillance System for Analyzing Abnormal Behavior of Pedestrians**. *Applied Sciences MDPI*, pp. 2-16.
6. Elsayed, O. A., Marzouky, N. A., Atef, E., & Salem, M. A.-M. (2019). **Abnormal Action detection in video surveillance**. *IEEE Ninth International Conference on Intelligent Computing and Information Systems (ICICIS)*.

7. Alhimale, L., Zedan, H., & Al-Bayatti, A. (2014, January 18). **The implementation of an intelligent and video-based fall detection system using a neural network**. Applied Soft Computing, pp. 59–69.
8. Cahoolessur, D., & Rajkumarsingh, B. (2020, June 12). **Fall Detection System using XGBoost and IoT**. R & D Journal of the South African Institution of Mechanical Engineering, pp. 8-18.
9. Li, H., Li, C., & Ding, Y. (2021, August 25). **Fall detection based on fused saliency maps**. Multimedia Tools and Applications, pp. 1883–1900.
10. World Health Organization (WHO). (2021, April 26). WHO. Retrieved 2 10, 2022, from <https://www.who.int/ar/news-room/fact-sheets/detail/falls>
11. Khraief, C., Benzarti, F., & Amiri, H. (2020, 3 25). **Elderly fall detection based on multi-stream deep convolutional networks**. Multimedia Tools and Applications, pp. 19537–19560.
12. Chaccour, K., Darazi, R., el Hassans, A. H., & Andres, E. (2015). **Smart Carpet using differential piezoresistive pressure sensors for elderly fall detection**. IEEE 11th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob).
13. Alwan, M., Rajendran, P. J., Kell, S., Mack, D. C., Dalal, S., Wolfe, M., & Felder, R. (2006). **A Smart and Passive Floor-Vibration Based Fall Detector for Elderly**. 2nd International Conference on Information & Communication Technologies.
14. Islam, Z. Z., Tazwar, S. M., Islam, M. Z., Serikawa, S., & Ahad, M. A. (2017). **Automatic Fall Detection System of Unsupervised Elderly People Using Smartphone**. 5th IIAE International Conference on Intelligent Systems and Image Processing. Japan.
15. RajuDhole, S., Kashyap, A., Dangwal, A. N., & Mohan, R. (2019, May 21). **A novel helmet design and implementation for drowsiness and fall detection of workers on-site using EEG and Random-Forest Classifier**. Procedia Computer Science, pp. 947-952.
16. Sarabia-Jácome, D., Usach, R., Palau, C. E., & Esteve, M. (2020, sep 26). **Highly-efficient fog-based deep learning AAL fall detection system**. Internet Of Things, pp. 1-19.
17. Thakur, N., & Han, C. Y. (2021, June 24). **A Study of Fall Detection in Assisted Living: Identifying and Improving the Optimal Machine Learning Method**. Sensor and Actuator Network, pp. 1-27.
18. Tahir, A., Ahmad, J., Morison, G., Larijani, H., Gibson, R. M., & Skelton, D. A. (2021, january). **HRNN4F:Hybrid Deep Random Neural Network For Multi-Channel Fall Activity Detection**. Probability in the Engineering and Informational Sciences, pp. 37-70.
19. Albawendi, S., Lotfi, A., Powell, H., & Appiah, K. (2018). **Video Based Fall Detection using Features of Motion, Shape and Histogram**. the 11th Pervasive Technologies Related to Assistive Environments Conference.
20. Sehairi, K., Chouireb, F., & Meunier, J. (2018). **Elderly fall detection system based on multiple shape features and motion analysis**. International Conference on Intelligent Systems and Computer Vision (ISCV). Fez, Morocco.
21. Ko, M., Kim, S., Kim, M., & Kim, K. (2018, june 15). **A Novel Approach for Outdoor Fall Detection Using**

- Multidimensional Features from A Single Camera.** applied sciences, pp. 1-22.
22. Thummala, J., & Pumrin, S. (2020). **Fall Detection using Motion History Image and Shape Deformation.** 8th International Electrical Engineering Congress (iEECON). Chiang Mai, Thailand.
23. Worrakulpanit, N., & Samanpiboon, P. (2014, December). **Human Fall Detection Using Standard Deviation of C-Motion Method.** Journal of Automation and Control Engineering, pp. 388-391.
24. Lakshm, K., & Devendran, T. (2020). **Human Fall Detection Using Motion History Image and SVM.** International Conference on Advanced Informatics for Computing Research.
25. Rougier, C., Meunier, J., St-Arnaud, A., & Rousseau, J. (2007). **Fall Detection from Human Shape and Motion History using Video Surveillance.** 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07).
26. frasetyo, M. B., Wahyuni, E. S., & Setiawan, H. (2019, April). **Comparison of Motion History Image and Approximated Ellipse Method in Human Fall Detection System.** IJCCS (Indonesian Journal of Computing and Cybernetics Systems), pp. 199-208.
27. Dubey, R., ni, B., & Moulin, P. (2012). **A Depth Camera Based Fall Recognition System for the Elderly.** 9th international conference on Image Analysis and Recognition - Volume Part II.
28. Yu, M., Gong, L., & Kollias, S. (2017, November). **Computer vision based fall detection by a convolutional neural network.** 19th ACM International Conference on Multimodal Interaction(ICMI).
29. Adhikari, K., Bouchachia, H., & Nait-Charif, H. (2017). **Activity recognition for indoor fall detection using convolutional neural network.** Fifteenth IAPR International Conference on Machine Vision Applications (MVA). Nagoya, Japan.
30. Anishchenko, L. (2018, may 7). **Machine learning in video surveillance for fall detection.** Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT), (pp. 99-102). Yekaterinburg, Russia.
31. Marcos, A. N., Azkune, G., & Carreras, I. A. (2017, Dec 6). **Vision-Based Fall Detection with Convolutional Neural Networks.** Wireless Communications and Mobile Computing, pp. 1-16.
32. Sangwan, D., Manekar, R., Saurav, S., Maiti, S., Singh, S., Chaudhury, S., . . . Chaudhary, K. (2018). **Activity Recognition for Indoor Fall Detection in 360-Degree Videos Using Deep Learning Techniques.** In Proceedings of 3rd International Conference on Computer Vision and Image Processing (pp. 417-429). New delhi, India: Springer Link.
33. Espinosa, R., Ponce, H., Gutiérrez, S., Villaseñor, L. M., Brieva, J., & Albor, E. M. (2019, November 27). **A vision-based approach for fall detection using multiple cameras and convolutional neural networks: A case study using the UP-Fall detection dataset.** Computers in Biology and Medicine.
34. Lie, W.-N., Hsu, F.-Y., & Hsu, Y. (2019). **Fall-down event detection for elderly based on motion history images and deep learning.** International Workshop on Advanced Image Technology (IWAIT) 2019. Singapore.
35. Khraief, C., Benzarti, F., & Amiri, H. (2019, Dec 19).

- Convolutional neural network based on dynamic motion and shape variations for elderly fall detection.** International Journal of Machine Learning and Computing, pp. 814-820.
36. Kong, Y., Huang, J., Huang, S., Wei, Z., & Wang, S. (2019, January 22). **Learning spatiotemporal representations for human fall detection in surveillance video.** Journal of Visual Communication and Image Representation, pp. 215-230.
37. Ramirez, H., Velastin, S. A., Meza, I., Fabregas, E., Makris, D., & Farias, G. (2021, February 23). **Fall Detection and Activity Recognition Using Human Skeleton Features.** IEEE Access, pp. 33532-33542.
38. Aldaqqaq, A., & Almahayri, M. (2021, July 27). **A Novel Approach to Detect Elderly Fall with Deep Learning Techniques.** Baath University Journal of Mechanical and electrical and Information (BUJMEI), pp. 1-28.
39. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). **ImageNet Classification with Deep Convolutional Neural Networks.** Advances in neural information processing systems 25, (pp. 1-9).
40. Davis, J. W. (1999). **Recognizing Movement using Motion Histograms.** Cambridge, USA: M.I.T Media Laboratory.
41. Dobhal, T., Shitole, V., Thomas, G., & Navada, G. (2015). **Human Activity Recognition using Binary Motion Image and Deep Learning.** Procedia Computer Science, pp. 178-185.
42. Tözeren, A. (2001). **Human Body Dynamics: Classical Mechanics and Human Movement. In Human Body Dynamics: Classical Mechanics and Human Movement** (p. 21). Washington, DC 20064, USA: Springer.
43. [Foroughi, H., Naseri, A., Saberi, A., & Yazdi, H. S. (2008). **An Eigenspace-Based Approach for Human Fall Detection using Integrated Time Motion Image and Neural Network.** 2008 9th International Conference on Signal Processing, (pp. 1499-1503).