Journal of Cybersecurity and Privacy

MDPI

Review

Deep Learning and Machine Learning, Better Together Than Apart: A Review on Biometrics Mobile Authentication

Sara Kokal [1], Mounika Vanamala [1,*] and Rushit Dave [2]

[1] Computer Science Department, University of Wisconsin, Eau Claire, WI 54701, USA; kokalsg4814@uwec.edu
[2] Computer Information Science Department, Minnesota State University at Mankato, Mankato, MN 56001, USA; rushit.dave@mnsu.edu
* Correspondence: vanamalm@uwec.edu

Abstract: Throughout the past several decades, mobile devices have evolved in capability and popularity at growing rates while improvement in security has fallen behind. As smartphones now hold mass quantities of sensitive information from millions of people around the world, addressing this gap in security is crucial. Recently, researchers have experimented with behavioral and physiological biometrics-based authentication to improve mobile device security. Continuing the previous work in this field, this study identifies popular dynamics in behavioral and physiological smartphone authentication and aims to provide a comprehensive review of their performance with various deep learning and machine learning algorithms. We found that utilizing hybrid schemes with deep learning features and deep learning/machine learning classification can improve authentication performance. Throughout this paper, the benefits, limitations, and recommendations for future work will be discussed.

Keywords: machine learning; machine learning algorithms; deep learning; behavioral biometrics; physiological biometrics

Citation: Kokal, S.; Vanamala, M.; Dave, R. Deep Learning and Machine Learning, Better Together Than Apart: A Review on Biometrics Mobile Authentication. *J. Cybersecur. Priv.* **2023**, *3*, 227–258. https://doi.org/10.3390/jcp3020013

Received: 12 April 2023
Revised: 6 June 2023
Accepted: 7 June 2023
Published: 13 June 2023

1. Introduction

Touch screen devices have evolved rapidly in recent years as demand and manufacture have skyrocketed. While smartphone capability continues to grow, progress in security has stagnated. This increasing gap between smartphone ability and security poses a significant problem. Today, smartphone users keep unprecedented amounts of sensitive data on their device, including photos, financial records, and private correspondence. Research into progressing the security of smartphone devices is necessary to protect the sensitive information of smartphone users.

To counter this issue, researchers and manufacturers have invested time and resources into developing and improving different types of smartphone authentication methods. Currently, the most common method of authentication in today's mobile phones are knowledge-based methods such as a password or a personal identification number (PIN) [1]. Since this method relies on the user's own knowledge, it runs the risk of the user choosing an easy to remember password that can easy be stolen or lost [1]. Due to the insecurity of knowledge-based authentication, researchers have turned to other methods such as physiological and behavioral biometrics. Physiological biometrics involve the unique physical characteristics of an individual such as their face, fingerprints, or iris. Behavioral biometrics involve how a person interacts with their device such as their typing, swiping, or tapping patterns [2]. Biometrics authentication applies the user's unique biological or behavioral features to phone security, which is more difficult to replicate by attackers in comparison to knowledge-based authentication [1]. In recent years, biometrics-based authentication methods have shown promising results when tested with machine learning and deep learning algorithms. Deep learning algorithm-centric methods have gained

popularity as of late due to recent tech advances enabling their efficiency, such as an increased availability of deep neural network (DNN) training datasets and increased computational power [3].

It has been found that biometrics-based authentication methods can more effectively secure mobile devices if more research is devoted to assuring that methodologies can be effective in real-world scenarios [4]. In this paper, the performance of machine learning and deep learning algorithms with biometrics-based authentication methods will be surveyed and analyzed. This subject has been investigated and reviewed by various researchers [4], thus this paper will continue this work by providing a comprehensive review on recent findings related to popular biometric methodologies with machine learning (ML) and deep learning (DL) algorithms. Our findings intend to help guide and inform researchers in their experiments relating to biometrics-based mobile security.

This paper is organized as follows: The background section is introduced in Section 2. It will discuss the differences between physiological and behavioral biometric dynamics, presenting examples as well as their individual benefits and limitations. Various types of DL and ML algorithms that will be spotlighted in the paper will be discussed and explained on their uses and structures. The bulk of this paper includes the literature review (Section 3), which has been divided into two main sections: behavioral and physiological dynamics. Behavioral biometrics constitute Section 3.1, and physiological constitutes Section 3.2. The behavioral biometrics section is further split up into different behavioral characteristics utilized for authentication: Touch (Section 3.1.1.), Motion (Section 3.1.2.), Keystroke (Section 3.1.3.), and Gait (Section 3.1.4.). The physiological section follows the same format, with Face (Section 3.2.1.), Ocular (Section 3.2.2.), and other (Section 3.2.3.). Following the literature review, we present the limitations of our analyzed studies (Section 4), a discussion of our findings (Section 5), a review of our research questions (Section 6), and conclude on what we believe are the most important findings for future work within the study of mobile biometric authentication (Sections 7 and 8).

This paper provides a comprehensive review of the most prevalent dynamics within biometric mobile authentication within recent and current research. As part of the process of this review, observations on trends with machine and deep learning algorithms are noted with regard to how they best perform with certain biometric dynamics. Other observations on the use of the datasets and data collection methodology are provided. This review aims to contribute to a better understanding of biometric mobile authentication with advice on future research projects based on patterns observed through our comprehensive analysis.

2. Background

Before proceeding with the literature review, it is important to provide an overview of the types of dynamics that will be discussed as well as some of the notable algorithms used with them. Physiological biometrics are effective since it is difficult to copy or share a unique physical characteristic. They are often performed as a method of static authentication [1]. One physiological dynamic that will be discussed in this paper is facial recognition, where the user is identified by matching captured images to the image stored in the device's database [1]. Facial recognition has benefits in authentication since faces are distinctive and usually readily available and unintrusive for capturing [5]. Nonetheless, facial recognition encounters challenges such as facial changes over time [5] as well as requiring high quality camera hardware that may not be up to par in all mobile devices [1]. Another dynamic prevalent in physiological biometric authentication is ocular recognition. The human eye contains many different features that can be used for authentication such as the iris and retina and has benefits due to the unchanging nature of the iris [5]. Challenges in ocular authentication include the difficulty of capturing the small retina, which requires specific hardware that may not be readily available in some mobile devices [1]. Fingerprint authentication is another popular physiological dynamic for mobile authentication and is used to secure a device by capturing and comparing fingerprint traits such as arches, loops, and whorls [1]. Fingerprint authentication is considered one of the most acceptable biomet-

rics today for user authentication [1]. While fingerprint recognition requires additional sensor hardware, fingerprint sensors are now more common in everyday devices, and are becoming cheaper and more accessible with time [5]. Other physiological dynamics that will be discussed in this review include voice authentication and vein recognition. Overall, physiological dynamics have benefits due to their unique nature to each individual, yet have the drawback of often requiring additional expensive hardware.

Behavioral biometrics have been called to attention for research since they can be captured without needing additional hardware or sensors [2] and can be used to dynamically authenticate while the device owner interacts with their phone [1]. Often, behavioral biometric authentication requires less direct user input than static methods. Behavioral biometrics-based models record how the user interacts with their device as data and uses it for authentication [1]. One dynamic that will be reviewed is touch-based authentication. Touch based authentication uses touchscreen inputs from the device, such as coordinates, pressure, and touch size to correctly identify the phone user. It is often paired with motion-dynamics to record phone micromovements while the device is swiped and tapped. Motion dynamics record data from motion sensors within most smart devices such as the accelerometer, gyroscope, and magnetometer. Motion data can be recorded from how the phone may be moved while in use. Another dynamic that will be discussed is keystroke authentication, which often involves a combination of touch and motion data. Keystroke dynamics use typing data to secure the device. The final behavioral dynamic that will be reviewed is gait dynamics. Gait dynamic-based methods record walking patterns using the phone's motion sensors. Gait dynamics are difficult to imitate, yet require movement to authenticate and are vulnerable to variation due to the user's environment [6]. While behavioral biometrics benefit as a dynamic and hardware-cheap method for authentication, they require events with the chosen trait to authenticate. If the user is not currently performing actions of the chosen trait, the behavioral system cannot secure the device, resulting in impractical time windows to detect intruders [6].

This review will discuss the performance of many different machine learning and deep learning systems with physiological and behavioral biometric models. ML and DL algorithms are a type of artificial intelligence (AI) that mimic human intelligence to make inferences on groups of data. DL algorithms are different from ML in that they make use of unsupervised learning strategies and can analyze unlabeled datasets [5]. Using DL strategies, algorithms learn hierarchical representations from large amounts of data with less human intervention. Prevalent DL algorithms in this review include convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep neural networks (DNNs). DNNs are a type of multilayer perceptron and utilize many hidden layers with fully connected weights in their architectures. RNNs and CNNs are both different types of DNN [6]. CNNs are a type of feed-forward network and are made up of stacks of layers that contain neurons with learnable weights and biases. These layers each transform 3D input volumes into 3D output volumes using a differentiable function [6]. RNNs can work with both supervised and unsupervised learning tasks and are commonly known for performing well with sequential data such as text. In unsupervised learning, RNNs use previous data samples to predict future data sequences, continuously updating and making predictions in a recurrent manner [6]. An example of a popular RNN used in many current studies would be a long short-term memory network (LSTM). Prevalent machine learning algorithms in this review include support vector machine (SVM) and random forest (RF). SVM is commonly used in classification and regression tasks. To perform classification, SVM constructs an n-dimensional space on which the datapoints lie and attempts to find the optimal hyperplane in the training dataset. The optimal hyperplane is a decision boundary that separates two classes of data and is used to properly classify new input data. RF is also used for classification and regression tasks. Unlike SVM, RF utilizes an ensemble learning technique. Multiple models, referred to as "trees", are constructed from the data samples and a majority vote is taken on their outputs. Both ML and DL

algorithms can provide acceptable classification performance depending on the needs of the authentication problem and each have their respective strengths and weaknesses.

3. Literature Review

3.1. Behavioral Biometrics

3.1.1. Touch Dynamics

Deep learning algorithms have been used to make classification decisions for many touch-based mobile authentication models. For studies [7–9], CNN architectures were used for classification. In [7], researchers used their proposed feature representation tactic, multiple channels biological graph (MCBG), with CNN to enhance their continuous mobile authentication scheme. MCBG takes touch and motion sensor data patterns and visualizes them into one of three feature graphs. These graphs are taken as input to train a CNN algorithm for classification. For model training and testing, a dataset of real-world touch interaction data from 180 participants was used. Their proposed model achieved a 96.77% accuracy with both moving and stationary data from the dataset. When compared with other classification algorithms, CNN had the best authentication accuracy with each feature graph input. Researchers of [8] evaluated implicit touch authentication with multidimensional touch and motion sensor data. For their classifier, CNN was used for its high performance with continuous implicit authentication (CIA) schemes. To use sequence data as CNN inputs, SE_ResNet with an attention mechanism was utilized to convert time series information into multichannel data. Datasets used for model training and testing included their own uncontrolled unlock pattern touch dataset with 50 users and the controlled BrainRun public dataset, consisting of various touch data collected from 2 k users via five interaction games. In the controlled scenario, an average accuracy of 96.94% was achieved, and in the uncontrolled scenario, a one-time recognition accuracy of 87.11% was achieved. These results provide evidence for the effectiveness of touch and motion sensor data with CIA. In [9], researchers analyzed deep models with touch dynamics authentication. For their model architecture, feature regularization net CNN was used to extract features from touch behavioral images, and a bidirectional LSTM was used to extract features from touch behavioral sequences. These features were then regularized and fused into a single layer before being sent to a K-nearest neighbor (KNN) algorithm for classification. For model training and evaluation, a multitouch gesture dataset was used, involving four-finger touch interaction data from 161 participants. The full version of their model, involving handcrafted feature concatenation and regularization and embedding loss, was able to achieve an equal error rate (EER) of 1.7%, outperforming comparison models.

Studies [10,11] chose LSTM as a classifier. Researchers of [10] analyzed a touch mobile authentication system with simple linear touch gestures. For their authentication model, a Siamese neural network architecture was used for classification that consisted of two LSTM layers. They trained their model with right swipe gesture data from the 600 subject HuMIdb database. From this touch data, they used 11 temporal features. They compared their model's performance with a Gaussian kernel binary-SVM. Results found that their model's performance saturated at six samples with 13% EER and outperformed the comparison SVM model. For [11], researchers presented their database MobileTouchDB and analyzed a passcode-drawing-based mobile authentication model. MobileTouchDB consisted of unsupervised touch data from 217 users assigned to draw numbers and characters on their device. A Siamese LSTM with dynamic time warping (DTW) was chosen for their classification model. During experimentation, the discriminative power of different characters as well as the effect of the length of passwords was evaluated. Their LSTM classification model was compared to four other systems. Results found that their time-aligned RNN achieved the best EER at 2.38%, and that EER decreased as password length increased until a length of four.

DNN architectures were used in studies [12,13]. In [12], scrolling and motion sensor-based touch data were evaluated with a DL model for two-class classification. A DNN consisting of three dense layers with either 64 or 128 nodes in each layer was used as a

classifier. Scrolling and motion sensor data from the Hand Movement, Orientation and Grasp (HMOG) dataset was used for training. Their model was evaluated on performance with different combinations of scrolling and motion sensor features. It was found that while more features did not guarantee better performance, pairing some motion sensor data with scroll data resulted in a better performance than scroll data alone. Gyroscope and scroll features combined resulted in the best performance with 80–81% average precision and 89–90% average accuracy. Other studies using the HMOG dataset are compared based on performance in Table 1. In [13], DL architecture was evaluated with their touch-stroke-based continuous authentication model, smart continuous authentication system (SCAS). The Touchalytics dataset containing touchscreen interaction and navigation data from 41 participants was used. Their DL architecture consisted of a first layer, which took portions of each stroke event as input data for their algorithms, and two hidden layers involving leaky rectified linear unit (Leaky ReLu) activation functions at each layer and SoftMax activation function at the output layer. Their model was able to achieve a testing accuracy of up to 94% with an average EER of 3.4%.

In studies [14,15], researchers chose multilayer perceptron (MLP) as a classification algorithm. Researchers of [14] evaluated a touch-and-motion sensor-based continuous authentication scheme. For their model, random forest was used to detect the motion state of the user, and MLP was used to authenticate the user with a model trained from their detected motion state. A cloud-based server was used to facilitate the registration and continuous authentication phased for their model. For model training and testing, the study used an unsupervised dataset consisting of touch and motion sensor data from 20 participants. Results found that their model with MLP classification was able to achieve an accuracy of 95.1% while using accelerometer and gyroscope data. MLP was able to outperform three other comparison algorithms. In [15], researchers proposed TouchMetric, an application designed to test machine learning algorithms with touch dynamics authentication. For their evaluation data, a thirty-four-subject dataset consisting of touch activity data from three categories (move activity, swipe activity, and type activity) was used to train and evaluate their models. Results from this study found that MLP was the only classifier of six to correctly classify both authenticated and unauthenticated users.

Studies [16–19] used other types of deep learning architectures and algorithms for classification. In [16], researchers proposed their touch-stroke-based two-class authentication model, kernel deep regression network (KDRN). Their proposed KDRN utilized stacking-based representation learning, consisting of multiple hierarchical layers of kernel ridge regression (KRR) trained analytically and independently. For training and testing, the Touchalytics dataset consisting of vertical and horizontal touch strokes from 41 participants was used. When compared with MLP, SVM, and radial basis function (RBF) kernel KRR, their KDRN model achieved the best results in all scenarios with an intra-session EER of $0.01 + -0.02\%$. KDRN produced the best EER performance between reviewed studies using the Touchalytics dataset as demonstrated in Table 2. For study [17], researchers proposed a continuous authentication scheme utilizing touch gesture dynamics. In this scheme, a radical basis function network (RBFN) with particle swarm optimization (PSO) was used for classification. PSO + RBFN was compared to four other classifiers, including decision tree (DT) J84, JRip, back propagation neural network (BPNN), and naïve Bayes (NB). To train their models, a dataset was collected from 20 participants tasked with performing various touch input such as single touch, multitouch, movement, and pattern-lock. PSO + RBFN was found to have the best performance with 2.0% FAR, 1.9% FRR, and 1.95% AER. In study [18], researchers proposed a touch stroke authentication model based on auxiliary classifier-generative adversarial network (AC-GAN). Their AC-GAN model consisted of a generator, discriminator, and a composite model with DNN structure in both the generator and discriminator. For model evaluation and training, the Touchalytics dataset was used. Authentication evaluation was performed by randomly selecting 10 subjects from the dataset to provide genuine user data, while the AC-GAN generated a balanced amount of synthetic imposter data to test against. Their model achieved EER, ranging from

2% to 11%, between subjects with a median of 7%, providing comparable performance to other touch stroke authentication systems. Researchers in study [19] proposed a continuous touch authentication system via a deep autoencoder with SoftMax regression (DAE-SR). For their model, butterfly optimization algorithm (BOA) was used to extract features before sending the input to DAE-SR for classification. Ten subjects' data from the HMOG dataset were used for model training and evaluation, where touch interaction and motion sensor data in controlled walking and sitting scenarios were collected. Their proposed model was able to outperform comparative classifiers SVM and KRR in both scenarios, with a best accuracy of 0.970 in sitting and a best EER of 0.030 in sitting.

Many other studies working with touch-based mobile authentication chose machine learning classifiers to make their authentication decisions. Studies [20–24] implemented SVM classifier in their models. In [20], researchers created an android app unlocking mechanism that utilized behavioral biometrics and ML for enhanced application security. Their application used pattern unlock to collect touch biometrics data and was evaluated via one class SVM (OC-SVM) and K-means ML algorithms. Their application and methodology were evaluated via 64 users and resulted in 93.7% true positive authorizations and 86% true negative authorizations. Touch biometrics data collected via pattern lock were also tested with OC-SVM in [21]. Researchers in this study aimed to address security issues with pattern lock via context awareness and fine-grained feature detection. In their model, motion sensor data collected from 77 users performing pattern lock input were used to identify posture context. Segmentation of pattern lock features via polylines enabled an improved accuracy and fine-grained detection. When this model was tested with different classifiers, OC-SVM outperformed. Their model was able to effectively improve pattern lock authentication with an EER of 5.1139% with context awareness and 98.96% accuracy with S pattern. In study [22], researchers tested SVM and Gaussian mixture model (GMM) with swiping biometric data from four different benchmark datasets: Serwadda, Frank, Antal, and the UMDAA-02 database. While SVM performed best with stable users, GMM performed best with unstable users. When fusing both SVM and GMM, the model produced better results overall in all sessions and datasets. Researchers found that the best performing operations were horizontal swipe gestures, and that landscape mode provided more stability. Their model with SVM and GMM fusion in intra-session scenarios was able to achieve 3–6% EER on the Serwadda dataset and around 3% EER with the Frank dataset. For [23], researchers Li et al. analyzed touch behavioral authentication when used with email application usage data. For their model, they collected touch action data from sixty participants in three scenarios: email usage, social networking usage and free usage. Five classifiers were compared on their performance while authenticating the collected data. A best average error rate (AER) of 2.9% was achieved by SVM classifier with email usage data. Researchers of [24] proposed a mobile authentication model via touch operations. Data from 10 subjects were collected while performing tasks such as tapping, swiping, and rotating. SVM classifier achieved better precision in single operations, but achieved a better accuracy when combining operations. The best accuracy for their model was 97.7% when tested with double tap and rotation data.

Studies [25–27] chose random forest as a classifier. In [25], researchers investigated the use of touch swipe and micromovement biometric data for authentication. For data collection, 40 subjects were tasked with answering a questionnaire using a slider. The data were divided into swipes and then extracted for features. For evaluation, many one-class and two-class classifiers were tested for user authentication. Results found that constrained horizontal swipes were effective with both one and two-class classifiers, and that accelerometer data improved performance. Random forest provided a best EER of (0.002 ± 0.000) with five swipes and eleven features. Researchers of [26] proposed a risk-based continuous authentication scheme with touch and motion data collected from unrestricted pin entry. Data were collected from 95 subjects in three scenarios: sitting, standing, and walking. These data were used to train a naive Bayes (NB) classifier, a neural network (NN), and a random forest classifier. Results found random forest to have

the best and most consistent results in all scenarios due to its ability to reduce variances and tune to overfitting. With 15 samples, RF achieved a highest TAR of 91.79%. For [27], researchers proposed an authentication scheme that identified hand motions (HM) and hold posture (HP) via touch and motion sensor interaction data for user verification. The study aimed to create a model that could authenticate in both dynamic and static scenes without constrained environments and interactions. For data collection, ten students posed as smartphone users and four researchers posed as imposters. KNN, SVM, and RF were trained and tested with the collected data, and all achieved above 94% accuracy, demonstrating the effectiveness of HMHP features. RF had the highest performing metrics with 0.99 AUC and >0.1% EER.

Researchers of [28–30] chose other ML classifiers for their models. In study [28], a continuous touch-dynamics authentication method was studied. Researchers used the Touchlogger application to collect background touch data during phone usage. For data collection, 14 participants collected general usage data over the course of two weeks, consisting mainly of single touch actions. RF, KNN, gradient boosting classifier (GBC), and linear SVM (L-SVM) classifiers were evaluated with these data. While RF and GBC had similar accuracy, GBC was chosen as the optimal classifier due to a faster performance with an average accuracy under the ROC curve (AUC) of 0.9692 and a learning time of 117 s. For [29], researchers attempted to create a privacy-preserving global continuous authentication model using touch dynamics. To achieve this, base features would be extracted from initial user sessions. Behavioral embedding was then performed, in which current session data would be transformed and embedded into past datasets. The global model would be trained on this embedded dataset. A total of 9 k subjects were tasked with data collection. Their method was tested on three algorithms: XGBoost (XGB), linear GBC, and DNN. The best results were achieved with L-GBC with an AUC of 0.913–0.921%, and an EER of 15.3–15.9%. Gradient boosting methods both outperformed the DNN. When compared with state-of-the-art (SOTA) techniques with OC-SVM and isolation forest (IF), their methods significantly outperformed. A touchscreen authentication model utilizing biometric information from both a smartphone and a smartwatch was proposed in [30]. For their model, a user performs a touch swiping task on the smartphone while wearing a smartwatch. The phone collects touch data while the watch collects motion data via accelerometer and gyroscope sensors. The scheme performs the authentication process using a server. Their server registers the user templates, stores data, and makes the authentication decision. The data flow for this authentication process is demonstrated in Figure 1. A Gaussian mixture model was used as a classification algorithm. To collect a dataset for their study, 20 volunteers were instructed to swipe on a Samsung device with three fingers in an L shape repeatedly for 120 times. Their model was found to perform with 91.5% authentication accuracy and resistance to attacks with only a 2.47% average attack success rate.

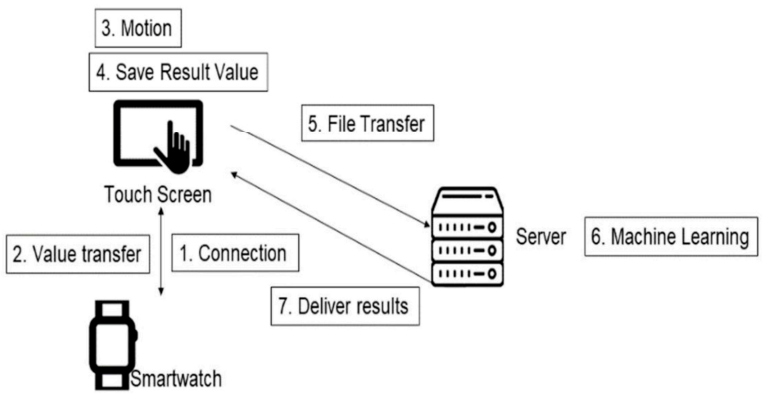

Figure 1. Data flow diagram for the authentication process [30].

3.1.2. Motion Dynamics

Deep learning algorithms are a popular choice for authentication in many recent studies involving motion dynamics. In studies [31–36], CNN was used to accomplish authentication. Hybrid models involving CNN feature extraction and SVM classification were noted frequently. Researchers of [31] aimed to combat DL vulnerability by testing a CNN model against adversarial attacks. Their authentication model was based on gesture authentication via the usage of time-series triaxial accelerometer data. To build their dataset, 16 users held a smartphone and traced a gesture in the air. A one-dimensional-CNN (1D-CNN) and an SVM were trained and tested on this data for authentication. Adversarial samples generated by deep convolutional GAN (DC-GAN) were used to evaluate CNN's robustness against adversarial attacks. Their CNN model slightly surpassed their SVM model with an accuracy range of 90% to 94% after hyperparameter fine-tuning was performed. Their CNN model was also able to resist adversarial attacks with a 90% accuracy facing poisoning attacks and almost completely overcoming evasion attacks. In study [32], researchers aimed to enhance the representational power of features and decrease the computational cost of continuous authentication models. Their model, SCANet, utilized a hybrid model involving a two-stream CNN based on depthwise separable convolution to extract features with high discriminability and an OC-SVM for classification. Their dataset included motion sensor data collected from 100 volunteers who performed tasks such as reading, writing, and navigation. When compared with other models, their CNN was able to achieve best accuracies of 87.53% and 90.04% in 2 s and 5 s time windows, respectively. A similar hybrid model and dataset were used in study [33]. Researchers proposed a continuous mobile authentication scheme using motion patters. A Siamese CNN performed feature extraction and OC-SVM performed classification. The HMOG dataset was used to evaluate their model. Their scheme produced a highest authentication accuracy of 97.8% with a sampling frequency of 25 Hz and a window size of 1 s in an analysis of all scenarios combined. Researchers of [34] also used a CNN–SVM hybrid model. Their approach generated grayscale images based on three-axis motion sensor signals from a single tap touch action and used these images to discriminate users. Touch data in the sitting position from the HMOG dataset were utilized for this study. A six-layer CNN model performed feature selection and an SVM was trained for classification. Their model was compared to LSTM feature and handcrafted feature models. Overall, their hybrid model significantly outperformed, with 96.72% accuracy. For study [35], a hybrid model for real-world mobile authentication was proposed. This model was aimed to combat issues with noise-in-motion sensor data when collected in real-world settings. A four-layer CNN was chosen to perform feature extraction and an SVM was used for classification. This CNN + SVM architecture is illustrated in Figure 2. A variational mode decomposition function (VMD) was used for data de-noising and signal enhancement. For their dataset, 1516 participants collected unlabeled and unsupervised motion sensor data over the course of one week. In comparison to other models using only ML, their CNN + SVM model outperformed with 95.01% accuracy. In [36], researchers proposed DeFFusion, a continuous mobile authentication scheme using motion sensor data. Their dataset consisted of time-domain accelerometer and gyroscope data collected from 100 users tasked to perform three common mobile phone tasks. These data were then converted into frequency data, from which a CNN extracts features and then fuses them to train an OC-SVM for classification. Results from this study found that when compared to OC-SVM, KNN, RF, and DT features, their CNN-based feature model performed the best with the SVM classifier. Their model produced the best authentication metrics of 1.00% EER, 1.42% false acceptance rate (FAR), and 0.75% false rejection rate (FRR).

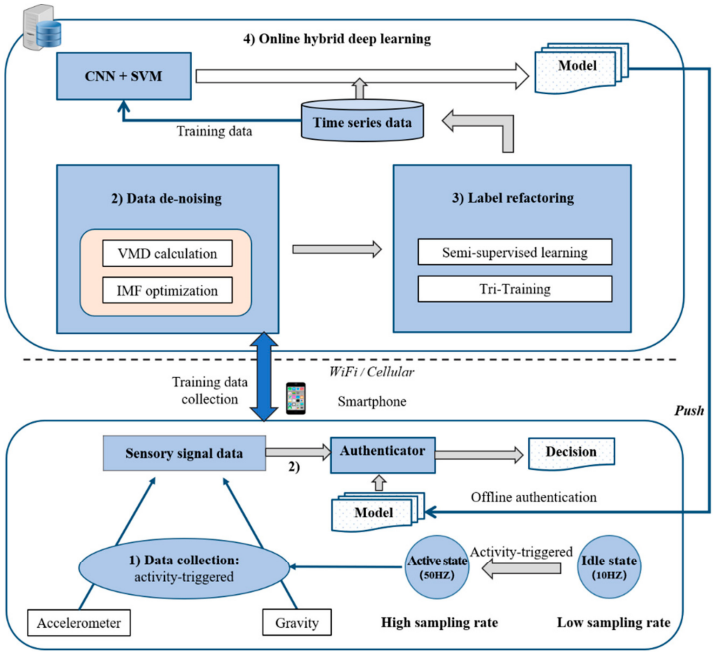

Figure 2. Proposed CNN + SVM architecture [35].

Studies [37–40] used RNNs such as LSTM for their authentication models. AutoSen, a continuous mobile authentication model utilizing motion sensor data, was proposed in [37]. For this model, data were captured implicitly without constraints on activity to preserve a realistic authentication scenario. Their dataset consisted of motion data from five smartphone sensors collected from 84 participants. An LSTM-RNN was used to capture behavioral traits directly from this dataset. Three different LSTM architectures were compared, including simple LSTM, bidirectional LSTM, and multilayer LSTM. Their analysis found that using data from only three sensors (accelerometer, gyroscope, and magnetometer) produced the best results. Multilayer LSTM was used for further authentication analysis and was able to achieve FRR 0.96%, FRR 8.08%, and EER 0.09% at 0.5 s frequency intervals. In study [38], DeepAuthen was proposed, utilizing motion sensor data for continuous authentication. Their model utilized DeepConvLSTM, a hybrid architecture consisting of a three-layer CNN tasked to handle spatial data and a one-layer LSTM to handle temporal data. Motion sensor data from the HMOG, UCI-HAR, and WISDM-HARB datasets were used to train and test their model. For evaluation, their CNN + LSTM model was compared to individual CNN and LSTM models. DeepConvLSTM outperformed the baseline models with a highest accuracy of 99.99% ($\pm$0.030%) and an EER of 0.01% ($\pm$0.060%) in the navigation during a sitting scenario from the HMOG dataset. Researchers of [39] proposed DeepAuth, an LSTM-based continuous authentication scheme using passive user behavior while online shopping. For their dataset, motion sensor data were collected from 47 users while browsing the Target application. Discrete Fourier transform (DTF) was then used to convert time signal data into frequency signals. An LSTM received these data for authentication. Their LSTM model was compared to four other baseline models: SVM, RF, logistic regression (LR), and GBC. Results found that DeepAuth outperformed the baselines and was able to detect both micro- and macro-movement patterns. Their highest metrics were 92.73% negative accuracy, 99.20% positive accuracy, 95.85% F1 score, and 99.05% AUC. In [40], researchers detailed a large-scale study on deep learning with human kinematics, proposing an active mobile authentication scheme via motion sensors. Their model used a specialized shift-invariant dense clockwork RNN for feature extraction and a GMM for classification. To test their model, researchers in this study created a dataset consisting of unrestricted motion sensor data collected from 1500 volunteers. When compared with other temporal models, their dense clockwork RNN produced the best accuracy with GMM.

The best metrics from their authentication evaluation included an EER of 18.17% and a human-targeted translation error rate (HTER) of 19.29%.

Table 1. Comparison of studies with HMOG dataset [12,33,34,38].

Study	Algorithm	Accuracy (%)
[12]	DNN	89.90
[33]	CNN + OC-SVM	97.80
[34]	CNN + SVM	96.72
[38]	LSTM	99.99

Table 2. Comparison of studies with Touchalytics dataset [13,16,18].

Study	Algorithm	EER (%)
[13]	DNN	3.4
[16]	KRR	0.01
[18]	AC-GAN	7.0

Researchers of [41,42] used other deep learning algorithms in their models. In [41], researchers evaluated a continuous mobile authentication model using deep learning autoencoders. Their scheme utilized a cloud system to manage communication between their classification algorithm and their collected data. For evaluation, this study used motion sensor data from two public datasets intended for continuous mobile authentication: the HMOG dataset and a crowdfunded dataset with data from 20 subjects. Their deep learning classifier with autoencoders was trained and tested on these datasets. Results found that their model was flexible enough to handle different authentication contexts, with an EER of 2.2% with five layers and 15 min intervals between model rebuilds. Researchers of [42] proposed a mobile authentication scheme utilizing the motion sensor data collected from the action of turning over a mobile device. In their dataset, 19 volunteers were instructed to turn their mobile phone over multiple times in different positions, from which accelerometer and gyroscope data were collected. Their classifier, a DNN, was trained with custom movement samples from this dataset. Results from this study found that their model produced viable metrics in both sitting and standing postures and was stable over time, with best metrics of 94–98% accuracy and 9–1% FRR and FAR.

Machine learning is also a popular method of classification for motion biometrics. Studies [43–46] used SVM classifier. In [43], researchers proposed a continuous authentication scheme via motion activity in different contexts. Their dataset consisted of motion sensor data from ten participants performing six different activities: walking, sitting, standing, running, and moving upstairs and downstairs. Three classifiers were tested with their authentication model: SVM, DT, and KNN. SVM achieved the highest accuracy results in all six activities between the comparison algorithms. SVM was decided to be the best for on-device user authentication with an overall average recognition accuracy of 97.95%. Researchers of [44] proposed RiskCog, a mobile authentication scheme using motion sensor data from both the mobile device and a smartwatch wearable device. Binary SVM with RBF kernel was used as a classification algorithm. Their model was tested with motion sensor data from six different public datasets. Results found that RiskCog was able to achieve higher accuracy rates than comparison studies, with 93.77% in steady conditions and 95.57% in moving conditions on dataset IV. Dataset IV consisted of data collection over 10 days from 1530 users. RiskCog was also able to resist brute force and mimicry attacks. Study [45] proposed SmartCAMPP, a continuous mobile authentication system via motion sensors. SmartCAMPP authenticates users by encrypting and preprocessing accelerometer and gyroscope data before authenticating with an ML classifier. Their model was tested on the Sherlock dataset, consisting of motion sensor data from 52 users over 500 h. Three different classifiers were tested, including SVM, RF, and LR. In baseline authentication results, SVM achieved a highest accuracy of 82.26%. SmartCAMMP's results produced an accuracy

of 76.85%, with a 5% reduction due to encryption. For [46], researchers analyzed behavioral biometrics to increase mobile authentication accuracy as compared to physiological. Their study analyzed a model based on accelerometer and gyroscope motion sensing. A dataset consisting of data recorded from 60 users lifting a phone to take a call was used to train and test four classifiers: SVM, KNN, MLP, and NB. Results found that while MLP had the highest average accuracy, it carried a heavy data processing footprint. SVM was claimed to be the best performing algorithm in both accuracy and efficiency with around 92% accuracy in both simple and complex scenarios.

In [47,48], researchers used RF as a classifier. For study [47], researchers proposed ADLAuth, an implicit mobile authentication scheme using motion sensor data from both the mobile phone and a wearable device. The methodology for their authentication system is shown in Figure 3. Their model used three different datasets ranging from 9 to 59 participants performing static and dynamic activities. SVM, DT, and RF classifiers were tested with the model and datasets. Results from the datasets only using smartphone sensing data achieved highest accuracy results with a random forest classifier. MobiAct dataset + RF classifier achieved the highest accuracy of 97.13%. In [48], AnswerAuth, a bimodal authentication scheme, was proposed. AnswerAuth used motion sensor data collected during the action of lifting a phone to take a call to discriminate users. Their dataset included accelerometer, gyroscope, magnetometer, and gravity sensor data collected from 85 participants in three different scenarios. Six different ML classifiers were tested with this model. Results found that RF outperformed the other classifiers with a highest accuracy of 99.35% with reduced features. Throughout the reviewed motion-sensing authentication schemes, it is observed that the accelerometer and gyroscope are of the most popular chosen motion sensors, as demonstrated in Figure 4.

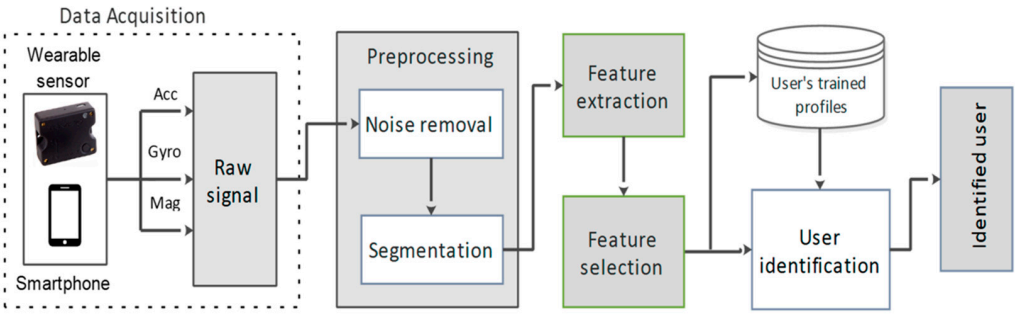

Figure 3. Proposed methodology for motion-sensing authentication [47].

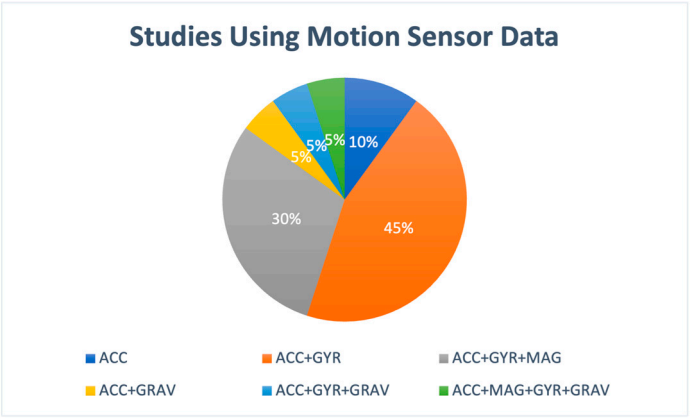

Figure 4. Proportion of motion-sensing studies using certain motion sensors for authentication. Accelerometer (ACC), gyroscope (GYR), magnetometer (MAG), and gravity sensor (GRAV).

3.1.3. Keystroke Dynamics

Many keystroke dynamics-based authentication methods have used deep learning algorithms to make their authentication decisions. Studies such as [49–54] used RNN-based deep architectures. In [49], researchers tested many types of touch modalities, including keystroke on their effectiveness with various motion sensors. They used HuMldb, a large-scale database including over 600 subjects. For the keystroke data, timestamp and keypress data were recorded in a fixed text scenario. Individual LSTM-RNN authentication models were trained with each modality and then fused at score level. Compared to other touch-based modalities, keystroke outperformed with a best EER of 4.62% when touch data were fused with accelerometer, magnetometer, and gravity sensor data. It was also found that the fusion of modalities led to a better performance. Similar results were found in [50], where researchers also used an LSTM-RNN model with Siamese architecture. Keystroke touch and temporal data along with motion and GPS sensor data were collected continuously in a free text scenario from a dataset of 37 subjects. Each modality was individually evaluated on their performance and fused. Their Siamese LSTM network was trained with deep temporal features of the various modality combinations. Results found that a highest TAR of 99.98% was achieved with all eight sensor modalities fused. Performance saturated after five modalities were used, with a 99.80% and 97.15% TAR at 0.1% FAR with 3 s and 5 s, respectively. In [51], researchers tested an RNN and other algorithms for mobile identification via keystroke and touch dynamics. The dataset used included keystroke and swipe data from 31 subjects collected in a supervised environment. Multimodal features were fused at score level and then used to train with various algorithms. RNN was able to achieve the best classification results with unimodal classification as well as multimodal classification. With all features fused late using weighted product rule, RNN achieved slightly better results than CNN with 95.29% accuracy and 1.78% EER. In [52], researchers proposed TypeNet, where an LSTM-RNN architecture was used for keystroke authentication. In the mobile scenario, free-text keystroke temporal data collected from 26,000 participants in the Palin et al. dataset was used for training an RNN with three loss functions. Their model was able to achieve 9.2% EER in mobile authentication with a triplet loss function when balancing enrollment data and performance. In [53], weighted results from and RNN and a SVM were used to authenticate keystroke data. A small dataset consisting of motion sensor, touchscreen, and temporal data from 10 users was used to train the models. It was found that a combination of motion sensor and touchscreen data had the best results, with 93.9% accuracy. Researchers of [54] proposed a touch database, BenchPassDB, which was evaluated with an LSTM-RNN classification model. BenchPassDB consisted of touch and motion sensor data from eight different touch interaction tasks performed by 81 subjects. Between their eight different touch tasks, keystroke interaction had the best authentication performance with 68.72% AUC. Their results found that discriminative ability was enhanced by a fusion of modalities.

Other methods [55–58] used non-RNN deep learning architectures. In [55], a CNN algorithm with transformer architecture was used for classification. The Aalto mobile keystroke database was used for model training and experimentation, consisting of free-text keystroke data from 260k subjects. Five timestamp-based features were extracted from each keypress. Results from their model were compared with the TypeNet [52] RNN model on the Aalto database. Their deep transformer architecture outperformed TypeNet on 10 enrollment sessions with 3.15% EER compared to 8.00% EER. For [56], a DNN classification model was created, consisting of layers of restricted Boltzmann machines (RBMs). The Stanford TapDynamics dataset was used for training and testing, consisting of tapping timestamp and accelerometer sensor data in a fixed PIN scenario. The results of the DNN model were compared with an SVM algorithm. Their model outperformed the SVM with a best EER of 2.8% on all 35 considered features. In [57], Kcollector was presented, using a multiview bagging with DL structures for classification. The DL structures used for the model consisted of gated recurrent unit multiview bagging RNNs (GRU-BRNN). Keystroke and accelerometer sensor data were collected from 26 subjects in a free-text scenario for

model training and experimentation. Compared to shallow machine learning approaches, their model outperformed with a best EER of 8.94% and 94.07% accuracy. In study [58], researchers evaluated a small dataset of five subjects with MLP for keystroke-enhanced PIN authentication. When compared with three other classifiers (sequential minimal optimization (SMO), naïve Bayes, J-48), MLP performed the best with all 40 pressure, size, and time features, achieving a 5.43% EER.

Some machine learning methods also achieved promising results with keystroke-authentication and were primarily tested with fixed keystroke data [59–62]. In [59,60], random forest was chosen as a classifier. Researchers of [59] used keystroke touch pressure and time features to enhance PIN authentication. Antal et al.'s 42 subject dataset consisting of fixed typing data was used to train a random forest classifier. Results found that random forest performed best while combining pressure and timing features with an EER of 2.3%. In [60], researchers used the same dataset as [59] to evaluate random forest. In this study, finger area was used as a feature, along with timing and pressure features, adding up to 71 features tested. Five statistical methods were then used on the dataset to produce 19 more features. Along with random forest, four other classifiers were evaluated with the dataset. Random forest outperformed the other classifier with 94.26% accuracy, using all 90 features. For study [61], a multiclass SVM was evaluated for a fixed-text keystroke authentication scenario. Researchers collected data in a fixed-text scenario from 94 subjects. Thirty-six optimal features involving touch pressure, size, and coordinates features were selected via maximum relevance minimum redundancy (mRMR) wrapping. Linear SVM and RBF kernel SVM were evaluated on these features. SVM with RBF kernel has the best results, with 97.4% accuracy due to its generalization power. In [62], researchers used PCA (principal component analysis) with keystroke data to enhance PIN authentication. Motion sensor data from accelerometer, gyroscope, and others were collected from 12 subjects while entering a PIN to form their dataset. A graphic visualization of each subject's data using PCA is shown in Figure 5. For evaluation, PCA was compared with Kernel-PCA (K-PCA), OC-SVM, and local outlier factor (LOF). PCA achieved the best performance with an EER of 5% with four-digit pins and 4% with six-digit pins. PCA also achieved the lowest processing time.

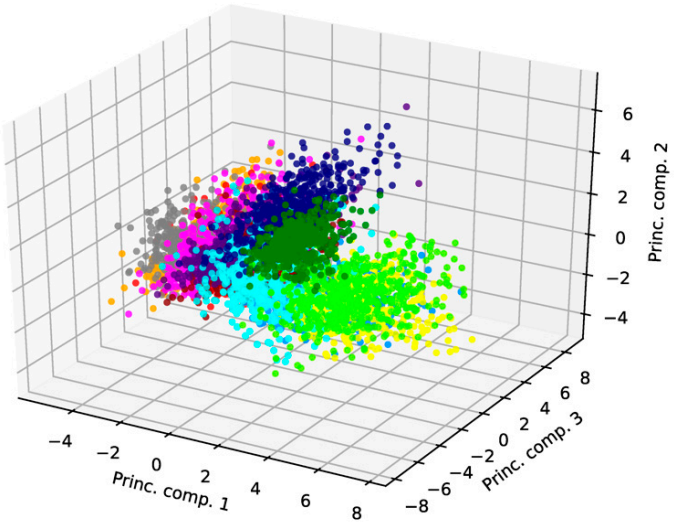

Figure 5. PCA visualization of the dataset. Each student is represented by a different color [62].

3.1.4. Gait Dynamics

Deep learning algorithms have been used in many studies on gait-based mobile authentication schemes. CNN-based models [63–65] have been used in gait dynamics authentication. IDNet, a gait mobile authentication scheme, was proposed in [63]. This model utilized a CNN for universal feature extraction and an OC-SVM for classification. Accelerometer and gyroscope data were collected from 50 subjects over a six-month period

to form their dataset. IDNet was built to be able to classify gait cycles regardless of smartphone orientation. For evaluation, CNN features were tested with multiple classifiers, and their IDNet was compared to other studies. Results found that all classifiers performed better with CNN features as opposed to manually selected features and that CNN features with SVM classifier had the best performance. Overall, IDNet outperformed the comparison literature with less than 0.15% FRR and FAR in fewer than five walking cycles. In [64], researchers explored the use of deep metric learning with gait-based mobile authentication. In this study, three CNN architectures (LeNet, VGG, and MobileNetv2) were compared based on their authentication performance with different Siamese networks (baseline, Siamese multiclass, Siamese binary-class). Four different datasets consisting of gait sensor data ranging from 20 users to 744 users were utilized to train the models. Results found that all three architectures had the best authentication performance with Siamese binary class and joint loss function. A best accuracy of 0.981% was achieved on the ZJU dataset with VGG8 architecture. For study [65], a model identification model using CNN and gait dynamics was proposed for smartphone-based sensing applications. This system extracts statistical and discrete Fourier transform (DFT) features from accelerometer data, which are then used to train a CNN model for identification. The Kaggle dataset, a real-world benchmark dataset consisting of accelerometer data collected from 387 users, was used in this study. For evaluation, their CNN model was compared to SVM, RF, DT, LR, and KNN. CNN performed the best on average with a highest accuracy (0.9882 ± 0.004), precision, recall, and F1.

Many studies [66–68] chose LSTM as a viable algorithm to pair with gait mobile authentication schemes. In study [66], a gait mobile authentication method utilizing three-axis accelerometer data was proposed. For this model, feature extraction was performed in the preprocessing stage. A hybrid deep recurrent neural network (DRNN) with a hidden LSTM layer was used for classification. A twenty-one-subject dataset consisting of flatland walking data in two smartphone-holding positions was used to train their model. For evaluation, the performance of their model was tested in different architecture parameters as well as between the two smartphone-holding positions (pocket and handholding). Results found that in handholding position, increasing the number of LSTM blocks resulted in a better classification rate and that two hidden layers was optimal for rate and evaluation time. For gait identification, their LSTM model performed better than RF. For gait authentication, their model was able to achieve a higher than 95% accuracy for almost all subjects in pocket position and 90% or higher in handholding position. Researchers of [67] continued their previous research on their smartphone imposter detection (SID) model, focusing on deep learning algorithms and protecting user data privacy. This study utilized the WALK section of the HAPT dataset, consisting of motion sensor data collected from 30 participants with smartphones strapped to their waists. Their model was evaluated in two scenarios: IDaaS, a binary classification scenario, and LAD, a scenario in which only user data are available for training. LSTM and SVM were compared in the LAD scenario and MLP was compared to SVM in the IDaaS scenario. DL algorithms achieved a better imposter detection than the ML algorithms in both scenarios with regard to a balance of execution time, memory usage, and accuracy. LSTM with prediction error distributions was able to achieve an accuracy of 90.24% in the LAD scenario with a 200 reading window. For [68], researchers proposed ContAuth, a continuous mobile authentication tested on gait, breathing, and electromyography (EMG) data. The architecture for this model used an LSTM for feature extraction and a stochastic gradient descendent (SGD) algorithm for classification. Incremental learning algorithms were also employed when encountering a new class. To train their model, four small-scale breathing datasets, an EMG dataset, and the IDNet [63] gait dataset were used in separate evaluations. With incremental learning and gait data, ContAuth was able to achieve an authentication accuracy of 97%, the highest between the dynamics tested. The performance of ContAuth is compared to other reviewed studies using the IDNet dataset in Table 3.Hybrid classification models using LSTM and CNN are the most popular choice for gait authentication schemes [69–73]. Researchers of [69] evaluated

the performance of mobile gait recognition in unconstrained conditions. Their hybrid DL architecture included a CNN for feature extraction and a two-layer LSTM for prediction. Accelerometer and gyroscope mobile sensor data from the WhuGAIT datasets were used to train and evaluate their model. This collection of datasets consists of mobile gait data collected from 118 participants in the wild. For evaluation, multiple combinations of DL methods were compared, and LSTM was evaluated with three different architectures: one layered, bidirectional, and two layered. Results from this study found that the accelerometer sensor better captured gait features when compared with gyroscope, but that they were complementary overall. LSTM and CNN were found to be effective with extracting inertial time series gait data. The hybrid CNN + LSTM model also outperformed standalone CNN and LSTM models, with a 93.75% authentication accuracy with vertically aligned samples. In study [70], an implicit gait mobile authentication model using edge computing was proposed. Their scheme involves generating their authentication model on a cloud server and deploying it on an edge device (mobile device) to preserve efficiency and computing resources. Their authentication model involves a CNN for feature extraction and an LSTM for classification, and the model was trained and tested using gait data from the IDNet dataset [63]. A visual diagram of the gait cycle is demonstrated in Figure 6. SVM and CNN were used as comparison baselines for model evaluation. The CNN–LSTM hybrid model outperformed with an accuracy of 97.7% and did not experience a significant dip in performance when less training data were incorporated. For [71], a cross-modal mobile authentication system via DL was evaluated. Their model consisted of a time-distributed cross-modal architecture with CNN feature extraction and LSTM interpretation. To train their model, researchers in this study used a fifty-person dataset consisting of walking data captured from accelerometer and gyroscope sensors. Their cross-modal CNN–LSTM model outperformed comparison algorithms (K-neighbors, DT, RF, CNN + PCA + OC-SVM) with a 92.3% accuracy. GaitPrivacyOn, a gait mobile authentication scheme, was proposed in [72]. This model utilized an architecture with two modules. In the first, an autoencoder is used to transform raw gait data to preserve user privacy. In the second module, a Siamese CNN–RNN is utilized for the model's verification system consisting of three convolutional layers and a bidirectional LSTM layer. Two public datasets consisting of motion sensor data from gait activity were used to train their model. Results found that this model was able to differentiate activity with 99.2% AUC and authenticate users with AUC ranging from 91.5% to 99.9%. In [73], researchers proposed a hybrid deep learning model (HDLN) for gait mobile authentication. Their hybrid architecture consisted of two LSTM layers and three one-dimensional CNN layers for feature extraction, followed by a convergence layer and a SoftMax layer for interpretation. A specialized segmentation algorithm was used to divide gait cycles to use as input for the HDLN. Accelerometer and gyroscope data from 40 individuals performing gait activity were used to train their model. A visualization of the gyroscope signals of three different subjects is presented in Figure 7. In comparison to other systems, HDLN performed the best with an accuracy of 95.79% and a run-time of 2.92 s. Using the IDNet dataset [63], this model was able to produce a 99.65% accuracy in a more realistic scenario.

Table 3. Comparison of studies on IDNet dataset [68,70,73].

Study	Algorithm	Accuracy (%)
[68]	LSTM + SGD	97.00
[70]	CNN + LSTM	97.70
[73]	CNN + LSTM	95.79

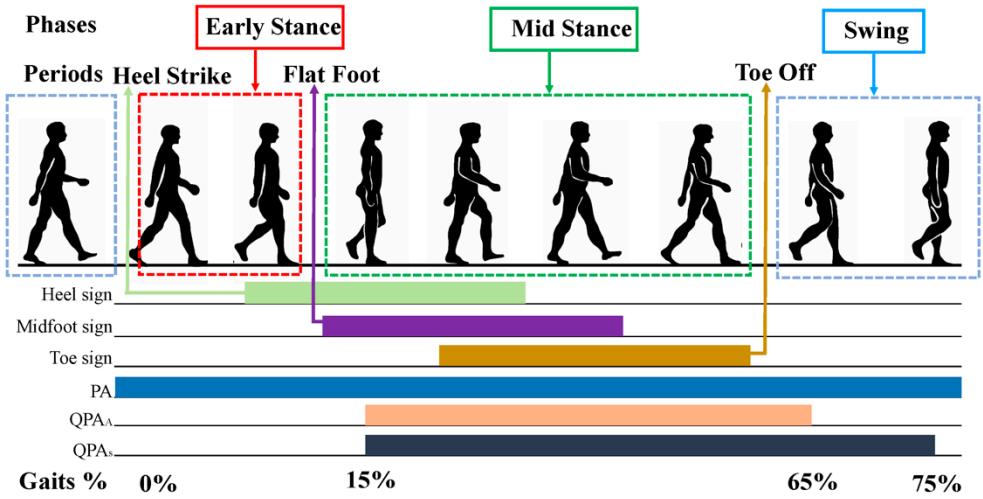

Figure 6. Depiction of a human gait cycle [70].

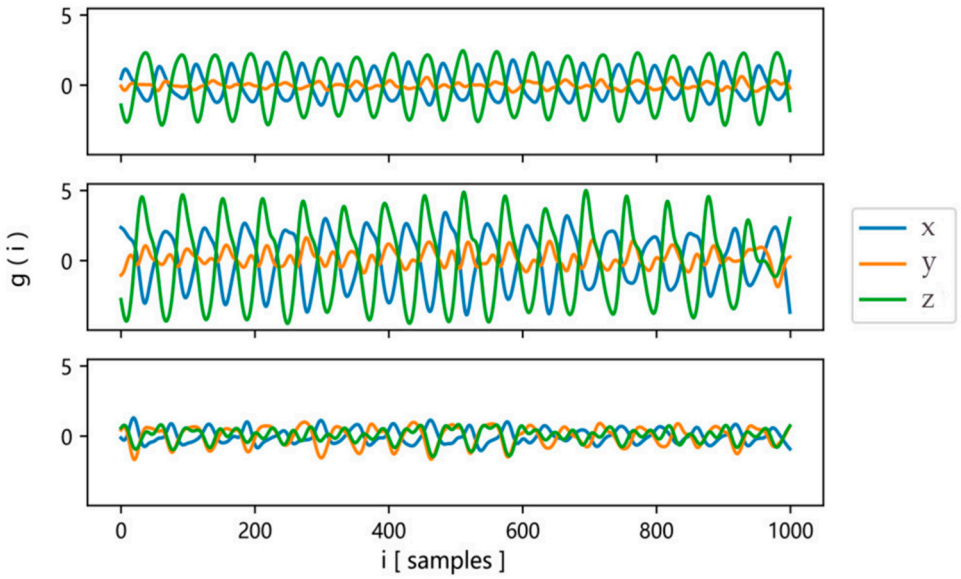

Figure 7. Gyroscope signals of three volunteers [73].

3.2. Physiological Biometrics

3.2.1. Facial Dynamics

Capturing facial features has been a popular method of mobile authentication in recent years. Many studies [64,74–77] have utilized the CNN algorithm for their authentication models. In [74], a cloud-based mobile face authentication system is proposed. On the user side, the device owner captures a face picture which is resized and sent to the cloud for authentication. On the cloud side, CNN is used for feature extraction and Softmax performs the classification. A total of 2800 facial images from 200 individuals in the FEI Face database was used for training and evaluation. The model achieved an accuracy of 99.50% and 0.01% loss. Efficient mask-net, a CNN-based facial authentication model, was proposed in [75]. The model architecture used a CNN for feature extraction and large margin piecewise linear classifier (LMPL) for classification. This system intended to authenticate users with and without face masks. To authenticate masked users, GAN was employed to generate a full face from the masked image, from which the system authenticates. Images from Masked Face-Net and Flicker Face-HQ datasets were used to train their model. Results found that their model was able to outperform other face-mask authentication models, with 99.53–99.64% accuracy with EfficientNetB0 + LMPL. In study [64], researchers address

the issue of face detection methods losing accuracy as the face distance is closer to the camera. They present a CNN-based mobile frontier face detection model. This method compared Viola Jones and CNN for face detection and alignment. The CNN FaceNet model was then employed for facial verification. This model was evaluated on a dataset of 40 individuals captured via a front-facing camera in three different distances. Results found that CNN outperformed Viola Jones for facial detection with a recognition accuracy of ~0.97. When compared with SOTA models such as PCA, latent Dirichlet allocation (LDA), Fisherface, and Eigen's face, CNN outperformed with ~0.99 verification accuracy. For [76], researchers present MobileFaceNets, a class of real-time CNN face verification systems. For their architectures, residual bottlenecks from MovileNetV2 are used as building blocks. PreLu is used as a nonlinearity. The primary MobileFaceNet has a large computational cost and 0.99 million parameters, so MobileFaceNet-M and MobileFaceNet-S were constructed as well. MobileFaceNet-M removes a linear 1×1 convolutional layer after the linear GDConv layer, and MobileFaceNet-S removes a linear 1×1 convolutional layer before the linear GDConv as well. The MobileFaceNets were compared to baselines such as MobileNetV, SuffleNet, and MobileNetV2. All models were trained on the CASIA-WebFace dataset and tested on the Labeled Faces in the Wild (LFW) and Age DB-30 datasets. The MobileFaceNets achieved an overall better performance compared to baseline studies, with a highest accuracy of 99.28% on the LFW dataset with the original MobileFaceNet. Models M and S were able to achieve extremely close results with less parameters and smaller computational cost. Researchers in study [77] propose a one class autoencoder regularized CNN (OC-ACNN) for active face authentication. The architecture consists of three nodes, feature extraction, classification, and decoder networks. In the feature extraction network, any CNN architecture can be used as the base, but AlexNet, VGGFace, and VGG16 were used for evaluation purposes. Before feeding input into the classification network, it is concatenated with a Gaussian vector to act as a pseudo-negative class. The classification network consists of a one-layer connected classification network. The decoder network consists of a four-layer CNN. The entire network is trained with two loss functions and three face datasets: MOBIO, UMDAA-01, and UMDAA-02. The proposed approach was compared with multiple baselines, including SVM and OC-NN. Overall, the OC-ACNN approach had superior performance compared to other OC classification methods on all three databases and all three extractor networks. VGGFace extractor achieved the best results with the OC-ACNN model on the UMDAA-01 database with 0.9772 ± 0.0213 AUROC.

While CNN alone can provide state-of-the-art mobile authentication, many studies have paired CNN with other ML algorithms in a hybrid model architecture [78–82]. Study [78] presents a multitask model for mobile authentication via facial attributes. Two CNN-based architectures are proposed: deep CNNAA and wide CNNAA. Both architectures consist of an ensemble of multitask DCNNs intended to extract facial attributes, with parameters varying depending on the facial region which is operated. Attributes extracted from the CNN architectures are then predicted by linear SVM. Two publicly available mobile-captured facial video and image datasets, MOBIO and AA01, are used for the active authentication evaluation problem. Results found that MutliDeep-CNNAA had better performance since the MultiWide-CNNAA had more parameters and overfitted to celebrity facial images. Overall, their attribute detection models performed better than comparison methods. In the active authentication problem, their method was able to outperform previous facial attribute studies. MutliDeep-CNNAA achieved an EER of 0.19–0.20% with the datasets tested. Between the reviewed studies using the MOBIO and A011 datasets, MutliDeep-CNNA presents the best EER performance as shown in Table 4. Echoprint, a mobile facial authentication system using both acoustic and visual features, was proposed in [79]. Echoprint utilizes the frontal camera to capture visual face features. For acoustic features, inaudible sound is emitted from the device to illuminate the physical face and capture features from the resulting echo. The architecture for this model included CNN feature extraction and SVM classification. Forty-five users supplied the data for their

training set, as well as five nonhuman images to provide sufficient data. Twelve users joined for model evaluation. Echoprint achieved a mean precision score of 98.05% when the 12 new users were evaluated on the model and were able to resist spoofing attacks. When comparing the usage of different features with different classifiers, it was found that SVM outperformed with CNN features, and that CNN features had the best accuracy rate overall. In [80], researchers trained and evaluated a small-sized mobile facial recognition model. Three deep learning solutions were compared for feature extraction in this dataset: FaceNet (CNN), OpenFace, and gb2s_model. Distance-based classifier (DBC) and SVM were compared as classifiers. To sufficiently train and test a model suited to scenarios as realistic as possible, a large sum of facial images was acquired from nine different public databases. Results found that DL solutions performed better than geometrical and local binary patterns (LBPs). The largest model, FaceNet, was able to achieve the best results with less variability between datasets as compared to the other models. When using SVM matching, system performance improved as compared to DBC. With SVM matching, FaceNet achieved an EER ranging from 0.00 to 2.56%. Ultimately, it is concluded that the feature extracting model directly influences results, and that the deeper the model, the better the results. Researchers of [81] created a model relying on facial recognition and eye-blinking biometrics to improve dynamic mobile authentication. Two phases make up their proposed model, with facial recognition performed in phase 1, and eye-blink authentication performed in phase 2. In phase 1, the face is detected via a histogram of oriented gradients (HOG) and linear SVM, from which facial landmarks are obtained. A CNN uses these landmarks to extract facial features, and an SVM performs classification. In phase 2, LeNet-5 CNN classifies the user's eye-blinking sequence. Datasets CEW and ZJU, consisting of eye photos and blinking videos, are used to train their eye-blinking classification model. Batch normalization was used to increase classification accuracy. Their model was compared with KNN and CNN + SVM models and outperformed with an FI score of 98.4% on the CEW dataset and an operation speed of 20 frames per second. In [82], researchers propose a mobile face authentication scheme via deep CNN with mobile cloud computing. The architecture of this model includes CNN + local ternary pattern (LTP) for feature extraction and KNN for classification. Facial images captured on the mobile device are stored on the cloud, which are preprocessed by LTP and sent to train the DCNN. The information is decrypted via partially homomorphic encryption (PHE), and the cloud makes the authentication decision, which is sent back to the device. Five facial datasets, including ORL, Yale Face database and the Georgia Tech Face database, are used to train and evaluate the model. The LTP-DCNN was able to achieve high performance in accuracy, precision, recall, and FI for all databases, outperforming existing methods. Recognition accuracy under encrypted data averaged 90.90–98.78% between the datasets.

Some studies utilized non-CNN architectures and algorithms for facial authentication and achieved comparable results [83–85]. A two-step face authentication scheme (TSFAS) was proposed in [83]. TSFAS utilizes a combination of facial biometrics and pin entry to enhance mobile banking security. The model begins once the user unlocks their device, with video feed monitoring their interactions. Model architecture involves facial feature extraction via denoised autoencoder (DAE). DAE is also used to perform the classification task. To train and evaluate TSFAS, the MSU-MFSD database was used, consisting of 280 video clips from 35 users as well as attack attempts. Results found that TSFAS was able to outperform two-faction authentication and hierarchical correlation-based authentication methods, with the best performing true rejection rates (TRRs) exceeding 92.58%. TSFAS was also able to resist spoofing attacks provided by the MSU-MFSD database. Study [84] proposed a generic prototype for a facial biometric mobile authentication model that preserves user privacy when working with services and transactions. Three components make up this model: the mobile device, the identity provider (IDP), and the service provider (SP). The IDP is the software that performs enrollment and authentication, while the service provider is the client application that facilitates the transaction. This protocol allows mobile users to authenticate to online services while preserving their privacy. The architecture

for the authentication model included a C-SVM for multiclass classification. To evaluate their model, the AT&T Laboratories Cambridge dataset was utilized. After many trails with different numbers of enrolled users in the authentication system and random users in the trained algorithm, their model was able to reach acceptable FAR and FRR levels. With a ratio of 1:2 enrolled and random users, the model was able to achieve an FAR of ~1%. In [85], researchers presented a continuous mobile authentication scheme with facial attributes biometrics. The PubFig database, consisting of facial images with labeled attributes, was used to train the attribute classifiers. This training process involves first detecting facial landmarks. Then, facial components are extracted. For each component, features are extracted and then dimensionally reduced with PCA. These features are used to train the classifiers. SVM with RBF kernel was chosen as a classifier and compared with many other methods using the MOBIO and AA01 mobile video face datasets. Linear-SVM on a 0.5 scale was able to achieve a best EER of 0.25%, outperforming the LBF method. The proposed model was implementable on a mobile scenario with acceptable memory usage, power consumption, and speed. Throughout the reviewed facial dynamics studies, CNN was the most popular chosen algorithm for architectures as demonstrated in Figure 8.

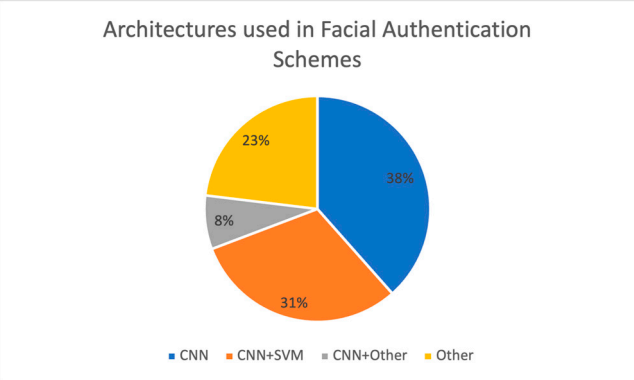

Figure 8. Figure depicting the proportion of architectures used in the cited facial authentication schemes.

Table 4. Comparison of studies with UMD-AA01 and MOBIO facial datasets [78,85].

Study	Algorithm	EER (%)
[78]	CNN + SVM	0.20
[85]	SVM	0.25

3.2.2. Ocular Dynamics

Ocular dynamics are a type of physiological biometric that have been investigated by researchers for their use in mobile authentication. Many studies [86–89] have focused on both ocular and periocular regions of an individual's eye to discriminate users. In study [86], researchers proposed a novel heterogeny-aware deep-embedding scheme for mobile periocular recognition. The architecture for this scheme involves a CNN in which the periocular images pass through. During training, a heterogeny-aware loss function optimizes feature representations to reduce intra-class variations. To train their model, three different datasets containing periocular imaging were used. Results found that their model outperformed comparison models with 99.41% accuracy and 1.32% EER with the VISOB dataset. Out of the reviewed studies using the VISOB dataset, this scheme produced the best EER values as shown in Table 5. OcularNet, a CNN-based ocular recognition model, was proposed in study [87]. This scheme utilizes a periocular region detected via eye landmark localization. Six overlapping patches are extracted from the periocular region, from which a CNN model is trained as a multiclass classifier for each patch. Four different datasets containing ocular imaging are used to train their models. Both ResNet

and OcularNet architectures are compared on their performance. Results found that despite being a smaller model, OcularNet was able to outperform ResNet, producing a lowest mean EER of 1.17% with the Oppo N1 device and the VISOB dataset. Researchers of [88], following the success of VISOB 1.0, present their findings from their most recent competition in mobile ocular biometric recognition. This publication first proposes their new VISOB 2.0 dataset, consisting of selfie image capturing of the periocular eye region from 150 volunteers for the training set and 100 for the testing set. For the selfie images, three sets of lighting were taken in burst mode, from which stacks of five images were used in the dataset. For the competition and evaluation, three teams participated. Team 1 used five ResNet-CNN models pretrained on the UGGFace dataset and fine-tuned with VISOB 2. Team 2 utilized decision tree–local binary patterns (DT–LBP), the only non-DL approach. Team 3 used GoogleNet extraction and LSTM prediction. Results from this study found that the deep learning approaches achieved better results. Specifically, the ResNet-CNN team achieved the best results with 5.256% EER and 0.988 AUC with Note4 challenge data. Researchers of [89] presented an in-depth study comparing multiple deep architectures for iris/ocular mobile authentication. Many CNN architectures such as VGG, ResNet, DenseNet, MobileNetV1, MobileNetV2 were compared with NasNet-Mobile and their own proposed architecture. Their proposed model was based off MobileNetV2, with a drop in spatial resolution and an increase in feature channels, resulting in a model with the least number of features and parameters. This model was modified to reduce computational complexity while retaining a mobile-capable authentication system. The VISOB dataset was reduced into three sections with around 200–400 participants each to suit their training and testing needs. Overall, it was found that ResNet-50, DenseNet-50, and the proposed model performed consistently better than the other comparison models. The study claimed their proposed model to be the best trade-off between performance and computational cost. In the subject-independent open set scenario, their model has the best EER values ranging from 4.65 to 6.57% EER.

Table 5. Comparison of studies with VISOB dataset [86,87,89].

Study	Algorithm	EER (%)
[86]	CNN	1.32
[87]	CNN	1.71
[89]	CNN	4.65

Other studies [90,91] using ocular biometrics for mobile authentication have utilized the many details of an individual's iris to perform user authentication. In study [90], researchers propose a multi-instance cancellable iris authentication system (MICBTDL). For the enrollment process, the user device captures left and right iris images. Random cross-folding is applied, and the images are transformed before being sent to a CNN for feature extraction. A reference template is generated and sent to a cloud server. The authentication process runs similarly, but the reference template is sent to the cloud server to be verified with an artificial neural network (ANN). To train their model, the MMU and IITD iris databases were utilized, consisting of left and right iris images of 253 subjects in total. Results found that their CNN extraction method outperformed comparison feature extraction techniques, and that MICBTDL achieved a fair performance against comparison studies, with an EER from 0.03 to 0.06% between the two datasets. Another CNN-based iris authentication system is proposed in study [91]. The architecture of their model involves a mask region-based CNN (R-CNN) structure that locates and extracts iris features with regional proposal networks component (RPN). For the recognition process, mobile inception V4 neural network architecture is fine-tuned and executed. The UTiris dataset, consisting of 1540 iris images captured from 79 users in non-constraint conditions, was used to train their model. Results found that their scheme was able to outperform comparison models with a 99.10% authentication accuracy. The study found their model suitable for high-performance mobile devices.

3.2.3. Other Physiological Dynamics

As previously mentioned, biometrics are currently the most prevalent in mobile biometric research, but they are not the only dynamics that have produced viable results for authentication schemes. Many studies such as [92–95] have utilized fingerprint characteristics for mobile authentication models. In [92], researchers proposed a partial fingerprint recognition scheme with DL. Zeng et al. recognized that modern mobile fingerprint scanners were not able to capture the entire fingerprint; thus, they created an algorithmic model built specifically for the recognition of partial fingerprint images. Their model used CNN-ResNet architecture with cross-entropy and contrastive loss functions. This model was trained using partial fingerprint images from the NCUT-FR and the NIST-DB4 databases, consisting of over 9 k fingerprint images in total, with 5 k collected from a touch fingerprint scanner. For evaluation, their model performance was compared to a reference study that also used CNN architecture for partial fingerprint recognition. Results found that the proposed model was able to outperform the reference model with 91.8–93% accuracy on the NIST-DB4 database. Study [93] proposed FINAUTH, a model aimed to complement fingerprint authentication models and resist puppet attacks with fingertip-touch characteristics. This model utilized sensors common in mobile devices such as accelerometer, gyroscope, and magnetometer to capture the fingertip-touch characteristics. For feature extraction, FINAUTH was tested with time-and-frequency domain features and CNN-learned features with Leaky-ReLu activation functions. For classification, the study compared the performance of four methods: Pearson's correlation coefficient (PCC), OC-SVM, LOF, and IF. For the collected dataset, 90 subjects and 15 attackers were recruited separately and tasked with performing fingerprint smartphone unlock in different holding positions. Results of their study found that CNN + LOF achieved the most reliable authentication performance and that CNN features were more discriminative than time–frequency domain features. In the sitting scenario, CNN + LOF achieved 97.99% BAC and 0.86% FAR. With non-overlapping subjects in training and testing, CNN-LOF achieved a 95.34% balanced accuracy (BAC) and 0.9805 AUC. Further testing with CNN-LOF found that FINAUTH was able to resist replica, puppet, and mimicry attacks. For study [94], researchers proposed a framework to detect spoofing in fingerprint recognition for mobile banking applications. Their model architecture first used histogram equalization on fingerprint images to increase contrast and uniformity before applying a CNN for feature extraction and classification. The CNN architecture included multiple convolutional layers and a two-unit SoftMax layer for the two-class problem (live vs. spoofed fingerprint). A total of five different databases consisting of real and spoofed fingerprint images was used to train their model. Results found that their CNN architecture was able to achieve >99.00% recognition accuracy on all databases, outperforming comparison SOTA CNN architectures. Spoof-detection rates were overall improved. C2CL, a contactless fingerprint matching mobile authentication system, was proposed in study [95]. For preprocessing, their method involved a U-net segmentation network, multiple image enhancements, and distortion correction and scaling performed by a spatial transformer network (STN). After preprocessing, a CNN architecture based on deep print was fine-tuned for the contact–contactless scenario and used to perform representation extraction. Finally, Verifinger 12.0 Software Development Kit (SDK) was used to extract minutiae representations from the contactless fingerprint images. A match score was computed with weighted fusion for classification/matching. Multiple databases consisting of over 50k contact/contactless fingerprint images were used for the training and testing of their model. Results found that C2CL was able to outperform previous SOTA verification results with significant improvement in EER values. On the PolyU database, C2CL achieved a 0.03% EER and 97.74 TAR @ FAR = 0.01%.

While few, other studies [96–98] have found voice and vein dynamics to be viable for mobile authentication. Following the COVID-19 epidemic, study [96] aimed to create a hygiene-aware contactless biometric recognition system utilizing wrist–vein biometrics. Their model was tested as a recognition system with a mobile application, utilizing wrist data captured via near-infrared camera with near-infrared lighting. The model architec-

ture consisted of a CNN algorithm for feature extraction and logistic regression (LR) for classification. CNN architectures such as VGG16, VGG19, ResNet50, and ResNet52 were compared in the evaluation stage. To train and test their models, the UC3M-SV2, UC3M-CV1, PUT, and ImageNet datasets were utilized. Results found that the models were able to produce viable EER rates in verification evaluation. The ResNet152 architecture achieved the lowest EER of 0.33% on the UC3M databases and the VGG16 architecture achieved a 0.58% EER on the raw image UC3M-CV2 database. Researchers of [97] presented a text-independent speaker authentication scheme for mobile devices. Their method involves recording voice audio and extracting features to create a voice sample dataset. Audio data are trimmed for voice segments and normalized, and a linear prediction cepstral coefficient (LPCC) is used for feature extraction. Voice signals before and after trimming are shown in Figure 9. Naïve Bayes (NB) was chosen as a classifier due to its low computational load. Three datasets were utilized for model training and testing, with a collected dataset of 11 student volunteers and two datasets from Ted talks and speakers (Ted-LIUM, TIMIT). With the proposed dataset, their scheme achieved 94% accuracy in a quiet environment and 83% accuracy in a noisy environment. A total of 83% accuracy was achieved with the Ted-LIVM dataset and 87% mean accuracy with TIMIT dataset. Results were found to be consistent with comparison schemes. ChestLive, an authentication system utilizing voice dynamics and chest movements, was proposed in [98]. While the user speaks, inaudible acoustic sensing is performed to capture reflected acoustic signals from the chest via the mobile speaker and microphone. These signals are then used to derive channel energy (CE) signals, from which interference is removed to produce a clear chest motion signal suitable for authentication. Mel-frequency cepstrum coefficient (MFCC) is used to extract features that suitably characterize both voice and motion signals. To address the problem of a small training sample set, meta-learning reptile algorithm is used for classification. In reptile, a general neutral network is first trained with a good generalization property and then adapted with new training samples. To construct their dataset, 61 volunteers were instructed to collect voice recordings in different usage scenarios with the mobile device pointed toward their chest. Results found that the model produced an average authentication accuracy of 98.31% and an EER of 2.92%. ChestLive was resilient in different usage scenarios and resistant to replay and impersonation attacks.

Figure 9. Voice signals before and after removal of silence segments [97].

We have observed from Table 6 that most published studies within behavioral biometrics are focused on the use of touch and motion sensor-based dynamics. Within physiological studies, a large majority invests research into facial and ocular dynamics. Across all studies, DL algorithms are shown as a popular choice for authentication models. Some recent studies have paired DL algorithms into hybrid architectures, performing as a feature extractor for an ML classifier or another DL classifier. It has been observed that many studies are taking advantage of an influx of public datasets containing carefully collected

data from large amounts of subjects, such as the HMOG, IDNet, MOBIO, VISOB, and HuMldb datasets. A great number of studies are achieving high metrics in accuracy and low metrics in EER values, demonstrating the potential for biometric mobile authentication security. Overall, the table represents a trend of high performance among studies evaluating biometric mobile authentication models.

Table 6. Comprehensive table of reviewed sources [7–98].

Dynamic	Study	Algorithms	Performance	Datasets
Touch	[7]	CNN	97.00% ACC	Original Dataset: 180 subjects
	[8]	CNN	96.94% ACC	Original Dataset: 50 subjects, BrainRun Dataset: 2218 subjects
	[9]	CNN + LSTM + KNN	1.7% EER	Outsourced Dataset: 161 subjects
	[10]	LSTM-RNN	13% EER	HuMldb Dataset: 600 subjects
	[11]	LSTM-RNN	2.38% EER	e-BioDigitDB: 93 subjects, MobileTouchDB: 217 subjects
	[12]	DNN	89–90% ACC	HMOG Dataset: 100 subjects
	[13]	DNN	3.4% EER	Touchalytics Dataset: 41 subjects
	[14]	MLP	95.1% ACC	Original Dataset: 20 subjects
	[15]	MLP	100% ACC	Original Dataset: 34 subjects
	[16]	KRR	95% ACC	Touchalytics Dataset: 41 subjects
	[17]	PSO + RBFN	1.95% AER	Original Dataset: 20 subjects
	[18]	AC-GAN	Median 7% EER	Touchalytics Dataset: 41 subjects
	[19]	DAE-SR	0.03% EER	HMOG Dataset: 10 subjects used
	[20]	OC-SVM + K-means	93.7% TPR	Original Dataset: 64 subjects
	[21]	OC-SVM	98.96% ACC	Original Dataset: 77 subjects
	[22]	SVM + GMM	3% EER	Serwadda, Frank, Antal, UMDAA-02 Datasets: 350 subjects total
	[23]	SVM	2.9% AER	Original Dataset: 10 participants
	[24]	SVM	97.7% ACC	Original Dataset: 10 subjects
	[25]	RF	0.002 ± 0.000 EER	Original Dataset: 40 subjects
	[26]	RF	91.79% TAR	Original Dataset: 95 subjects
	[27]	RF	>0.1% EER	Original Dataset: 14 subjects
	[28]	GBC	0.9692 AUC	Original Dataset: 14 subjects
	[29]	L-GBC	0.913–0.921% AUC	Original Dataset: 9 k subjects
	[30]	GMM	91.5% ACC	Original Dataset: 20 subjects
Motion	[31]	CNN	90–94% ACC	Original Dataset: 16 subjects
	[32]	CNN + OC-SVM	90.04% ACC	Original Dataset: 100 subjects
	[33]	CNN + OC-SVM	97.8% ACC	HMOG Dataset: 100 subjects
	[34]	CNN + SVM	96.72% ACC	HMOG Dataset: 100 subjects
	[35]	CNN + SVM	95.01% ACC	Original Dataset: 1513 subjects
	[36]	CNN + OC-SVM	1.00% EER	Original Dataset: 100 subjects
	[37]	LSTM	0.09% EER	Original Dataset: 84 subjects
	[38]	LSTM	0.01% EER	HMOG Dataset: 100 subjects, UCI-HAR Dataset: 30 subjects, WISDM-HARB Dataset: 51 subjects
	[39]	LSTM-RNN	99.05% AUC	Original Dataset: 41 subjects
	[40]	RNN + GMM	18.17% EER	Original Dataset: 1.5 k subjects
	[41]	DAE	2.2% EER	HMOG Dataset: 100 subjects, Original Dataset: 20 subjects
	[42]	DNN	94–98% ACC	Original Dataset: 19 subjects
	[43]	SVM	97.95% ACC	Original Dataset: 10 subjects
	[44]	SVM	95.57% ACC	Outsourced Dataset IV: 1513 subjects
	[45]	SVM	76.85% ACC	Sherlock Dataset: 52 subjects
	[46]	SVM	92.0% ACC	Original Dataset: 60 subjects
	[47]	RF	97.13% ACC	HAR Dataset: 30 subjects, PAMAP2 Dataset: 9 subjects, MobiAct Dataset: 59 subjects
	[48]	RF	99.35% ACC	Original Dataset: 85 subjects

Table 6. *Cont.*

Dynamic	Study	Algorithms	Performance	Datasets
Keystroke	[49]	LSTM-RNN	4.62% EER	HuMldb Dataset: 600 subjects
	[50]	LSTM-RNN	99.98% TAR	Original Dataset: 37 subjects
	[51]	RNN	1.78% EER	Original Dataset: 31 subjects
	[52]	RNN	9.2% EER	Palin et al. Dataset: 260 k subjects
	[53]	RNN + SVM	93.9% ACC	Original Dataset: 10 subjects
	[54]	LSTM-RNN	68.72% AUC	BehavePassDB: 81 subjects
	[55]	CNN	3.15% EER	Aalto Dataset: 260 k subjects
	[56]	DNN	2.8% EER	Stanford TapDynamics Dataset: 55 subjects
	[57]	GRU-BRNN	94.07% ACC	Original Dataset: 26 subjects
	[58]	MLP	5.43% EER	Original Dataset: 5 subjects
	[59]	RF	2.3% EER	Antal et al. Dataset: 42 subjects
	[60]	RF	94.26% ACC	Antal et al. Dataset: 42 subjects
	[61]	L-SVM + RBF	97.4% ACC	Original Dataset: 94 subjects
	[62]	PCA	5% EER	Original Dataset: 12 subjects
Gait	[63]	CNN + OC-SVM	<0.15% FRR, FAR	Original Dataset: 50 subjects
	[64]	CNN (VGG8)	0.981 ACC	Mcgill: 20 subjects, IDnet: 50 subjects, ZJU: 153 subjects, Osaka: 744 subjects
	[65]	CNN	0.9882 ± 0.004 ACC	Kaggle Dataset: 387 subjects
	[66]	DRNN-LSTM	>95% ACC	Original Dataset: 21 subjects
	[67]	LSTM	90.24% ACC	HAPT Dataset: 30 subjects
	[68]	LSTM + SGD	97% ACC	IDNet Dataset: 50 subjects
	[69]	CNN + LSTM	93.75% ACC	WhuGAIT: 118 subjects
	[70]	CNN + LSTM	97.7% ACC	IDNet Dataset: 50 subjects
	[71]	CNN + LSTM	92.3% ACC	Outsourced Dataset: 50 subjects
	[72]	CNN + LSTM	91.5–99.9% AUC	MotionSense Dataset: 24 subjects, MobiAct Dataset: 56 subjects
	[73]	CNN + LSTM	95.79% ACC	IDNet Dataset: 50 subjects
Face	[74]	CNN	99.50% ACC	FEI Dataset: 200 subjects
	[75]	CNN + LMPL	99.53–99.64% ACC	Masked FaceNet, Flicker Face-HQ Datasets: 15 k images
	[64]	CNN	~0.99 ACC	Original Dataset: 40 subjects
	[76]	CNN	99.28% ACC	CASIA-Webface, LFW, AgeDB-30 Datasets: unspecified subjects
	[77]	OC-ACNN	0.9772 AUROC	MOBIO Dataset: 150 subjects, UMDAA-01 Dataset: 50 subjects, UMDAA-02 Dataset: 44 subjects
	[78]	CNN + SVM	0.19–0.20% EER	MOBIO Dataset: 150 subjects, UMDAA-01 Dataset: 50 subjects
	[79]	CNN + SVM	98.05% Precision	Original Dataset: 57 subjects
	[80]	CNN + SVM	0.00–2.56% EER	BioID, EUCFI, ORL, Ext. Yale B, PrintAttack, gb2sTablet, gb2sMOD, gb2s_Selfies, gb2s_IDCards: 696 subjects
	[81]	CNN + SVM	98.4% FI-score	CEW Dataset: 2423 images, ZJU Dataset: 80 video clips
	[82]	CNN + KNN	90.90–98.78% ACC	ORL: 40 subjects, Yale: 15 subjects, Extended Yale: 40 subjects, Georgia Tech: 50 subjects, FEI: 200 subjects
	[83]	DAE	>92.58% TRR	MSU-MFSD Dataset: 35 subjects
	[84]	SVM	~1% FAR	AT&T Dataset: 40 subjects
	[85]	SVM	0.25% ERR	MOBIO Dataset: 150 subjects, UMDAA-01 Dataset: 50 subjects

Table 6. *Cont.*

Dynamic	Study	Algorithms	Performance	Datasets
Ocular	[86]	CNN	99.41% ACC	VISOB Dataset: 550 subjects, CSIP Dataset: 50 subjects, IIITD Dataset: 62 subjects
	[87]	CNN	1.17% EER	VISOB Dataset: 550 subjects, UBIRIS-1 Dataset: 241 subjects, UBIRIS-2 Dataset: 261 subjects, CrossEyed Dataset: 120 subjects
	[88]	CNN	0.988 AUC	VISOB-2 Dataset; 150 subjects
	[89]	CNN	4.65–6.57% EER	VISOB Dataset: 550 subjects
	[90]	CNN + ANN	0.03–0.06% EER	IITD Dataset: 225 subjects, MMU Dataset: 45 subjects
	[91]	CNN	99.10% ACC	UTiris Dataset: 79 subjects
Other	[92]	CNN	91.8–93% ACC	NCUT-FR Dataset: 5 k images, MST-DB4 Dataset: 4 k images
	[93]	CNN + LOF	97.99% BAC	Original Dataset: 105 subjects
	[94]	CNN	>99.00% ACC	FVS2006, ATVSFFpDB, Spoof Attack Finger Vein Database, LivDet 2013 Dataset, LivDet 2015 Dataset
	[95]	CNN	0.03% EER	UWA Benchmark, ManTech Phase2, PolyU, MSU, IIT, ISPFDv2, ZJU Datasets: ~1 k subjects total
	[96]	CNN	0.33% EER	UC3M-CV2: 2400 images, UC3M-CV1: 1200 images, PUT: 1200 images, ImageNet: 14 million images
	[97]	Naïve-Bayes	94% ACC	Original Dataset: 11 subjects
	[98]	NN	98.31% ACC	Original Dataset: 61 subjects

4. Limitations

Despite the promising results and insightful contributions of the various studies presented in this paper, many have limitations in their data collection and research conduction that could affect the legitimacy and relevancy of their proposed models. One common limitation occurring in many studies reviewed would be the usage of small datasets. While it can be difficult to accumulate or find datasets relevant to a study with large amounts of data, it is significantly important to the quality of modern authentication schemes. As DL becomes more relevant in the study of biometrics authentication, the need for larger datasets has increased. DL algorithms require a substantial amount of data to adequately train their models in comparison to ML models. Small datasets also affect data quality in that significant data variability can lead to harmful bias and underrepresentation of the target population. This results in a model that cannot accurately classify the target population. Another frequent limitation would be a lack of testing against attacks. While some studies such as [30,31,44,79,83,93,98] were engaged in testing their models against attacks, a vast majority did not. There are many different types of attacks designed to combat various types of mobile authentication strategies, such as smudge, over-the-shoulder, replica, puppet, mimicry, replay, and impersonation attacks. As mobile authentication schemes become more advanced, it is important to make sure our models can effectively resist these types of attacks. Another limitation prevalent in many studies would be constrained scenarios in data collection and authentication testing. In the cases of many behavioral biometric schemes such as gait, touch, and motion dynamics, there can be a lot of variability in how the individual uses their device and performs an activity. Often, these data are difficult to process for authentication. While some studies such as [7,8,35,65] account for this variability in their data collection and preprocessing stages and allow unconstrained data collection and realistic smartphone usage, many studies instruct their volunteers to follow more rigid instructions for data collection and testing. This can result in models

that are not equipped to handle authentication in real-life scenarios. If biometrics mobile authentication is to become a practical option for mobile phone security, these limitations in our studies need to be addressed.

5. Research Questions

We have aimed this review to provide answers to questions related to the use of ML and DL algorithms with biometric authentication. Question 1: Which biometrics are the most effective between behavioral and physiological biometric mobile authentication? We have observed that touch and motion dynamics are the most prevalent and effective within behavioral biometrics, while facial and ocular dynamics are largely focused on within physiological dynamics. Question 2: What algorithms are effective with regard to biometric dynamics in mobile authentication? This has been answered through our observations and conclusions from Table 6. Question 3: What types of algorithms should researchers investigate for their studies on biometric mobile authentication? We have concluded from our review that deep learning algorithms are especially effective with biometric mobile authentication, and that combining two AI algorithms in a hybrid scheme can provide better performance than a single algorithm alone.

6. Discussion

Prevalent DL and ML algorithms and their usage with various biometric authentication schemes have been reviewed in depth throughout this paper. After close analysis of each biometric dynamic, it is clear that the careful selection of algorithms can result in better performance, and that specific algorithms are more adept to working with different types of biometric data. Considering deep learning authentication schemes, CNN and RNN dominate physiological and behavioral dynamics. Both are proficient with touch, motion, and gait dynamics, as demonstrated in Table 6. RNN is observed to be especially useful with keystroke dynamics, which was anticipated given that RNN is known to perform well with sequential data such as text. CNN was recognized to be proficient with physiological data, performing with higher metrics compared to other algorithms with facial, ocular, and fingerprint-based authentication schemes. This trend is consistent with CNN, given that CNN is majorly used in image recognition, and physiological dynamics commonly utilize image-captured inputs for their authentication schemes. Overall, CNN and RNN are the two most used algorithms within the studies examined within this paper as seen in Table 6 and can produce authentication results with a high accuracy and capability with mobile devices. With the rise of dataset quantity and quality in recent years, DL algorithms have been able to surpass ML in many studies and prompt the creation of more advanced and capable mobile authentication models.

As previously established in [4], SVM is a strong contestant for mobile behavioral biometric classification. This was once again observed within studies in this paper. SVM was popular in studies working with touch, motion, and keystroke dynamics schemes, outperforming many other ML algorithms. RF was found to be a close second, also providing adequate performance with these schemes with favor to keystroke dynamics. It is notable that SVM was also observed to be capable as a classifier in physiological dynamics such as facial authentication when paired with CNN features. This brings the discussion of a particular phenomenon within mobile biometric authentication to light.

Hybrid authentication systems are on the rise, as demonstrated by many models within this paper. Many schemes using hybrid architectures involve using one algorithm for feature extraction and another to perform classification with those features. These architectures have presented themselves in pairs of DL–DL as well as DL–ML. CNN was exceedingly popular as a feature extractor for many of these hybrid authentication systems. Within gait dynamics, it was observed that a CNN + LSTM system was the most popular choice out of any architecture, with CNN used for feature extraction and LSTM for classification. CNN + LSTM often outperformed standalone LSTM and CNN models as demonstrated by [69–71]. Within facial authentication, it was observed that a CNN + SVM

architecture was prevalent, wherein CNN was used for feature extraction and SVM was used for classification. In studies [79–81], different combinations of features with different classifiers were analyzed with facial authentication. In these studies, CNN features with SVM classification outperformed other combinations. Motion-sensing dynamics authentication systems in studies [32–36] also used CNN + SVM hybrid architectures. SVM was found to be the best performing classifier with CNN features in studies [32,36]. In study [34], using CNN features with SVM performed better than with handcrafted features or LSTM-based features. Comparing their hybrid model with non-hybrid comparison studies, researchers of [35] found their CNN + SVM to have a better accuracy. When analyzing these trends, it is evident that hybrid architectures are promising, with mobile biometric authentication able to enhance the authentication performances of many classifiers with CNN features.

7. Conclusions

This survey has presented a comprehensive review of the current state of mobile biometric authentication, highlighting the most significant algorithms and popular dynamics. Within behavioral biometric authentication schemes, touch and motion dynamics were popular choices for mobile authentication models, with the greatest number of studies present out of all sources. Touch dynamics schemes performed well with a variety of DL and ML algorithms, mainly featuring CNNs, RNNs, and SVMs. Motion dynamics schemes were also effective with CNNs and SVMs, especially when paired together to form a hybrid CNN + SVM architecture. Within motion schemes, accelerometer and gyroscope were prevalent, as demonstrated in Figure 4. In keystroke dynamics models, LSTM–RNNs were outstandingly common and proficient. Schemes that utilized gait dynamics worked well with CNNs and RNNs, specifically when combined in a hybrid CNN + LSTM architecture. In physiological authentication models, CNN completely dominated. As demonstrated in Figure 8, CNN architectures were used in the majority of facial authentication models, providing sufficient performance both as a standalone classifier and a feature extractor. This trend persisted in other physiological models using ocular, fingerprint, vein, and voice characteristics.

To summarize, deep learning has been found to be an effective method of classification in a variety of different contexts within behavioral and physiological authentication. CNNs and RNNs were overwhelmingly the most effective and popular algorithms for mobile biometric authentication schemes presented in studies from the past five or so years. CNN has proven to be exceptional not only in authentication contexts across both behavioral and physiological dynamics, but also as a feature extractor capable of enhancing the classification performance of many other algorithms. Most notably, hybrid authentication schemes have been observed to be a compelling and advancing architecture for classification models, resulting in higher accuracies than standalone models in many studies such as [35,67,69–71].

Overall, the usage of DL algorithms and hybrid authentication models is worthy of and in need of further investigation within the field of mobile biometric authentication. The rise in the prevalence of DL algorithms across both behavioral and physiological dynamics indicates great change in the field of mobile biometric authentication, as studies investigate more complex architectures for their models. With many studies starting to take advantage of hybrid architectures, it is reasonable to assume that hybrid DL architectures have the potential to advance biometric authentication past previous boundaries. While behavioral biometric authentication requires more research and experimentation to become usable in real-world scenarios, experimentation in hybrid architectures shows that growth in this field is possible and that investment into the study of biometric mobile authentication is worthwhile.

8. Future Work

For future work, researchers should look toward deep learning algorithms such as CNNs and RNNs for their biometric mobile recognition systems. Exploring DL algorithms

as well as their potential in hybrid authentication schemes is highly encouraged due to their observed potential in the schemes reviewed. Figure 10 visualizes the hybrid architecture strategy currently used in newer studies, which is projected to enhance the performance of biometric authentication with continued investigation. Furthermore, addressing limitations within dataset and data collection quality as well as realism in testing scenarios is necessary to ensure that authentication schemes are suitable for real-world contexts. With the rise in DL algorithms, it is important that researchers focus especially on expanding their datasets in both quality and numbers to enable the full potential of DL and create an ideal biometric authentication model. This ideal model is depicted in Figures 11 and 12 for behavioral and physiological dynamic authentication. Further research in mobile biometrics authentication is encouraged to engage in study in the aforenoted areas and address the current limitations as previously stated.

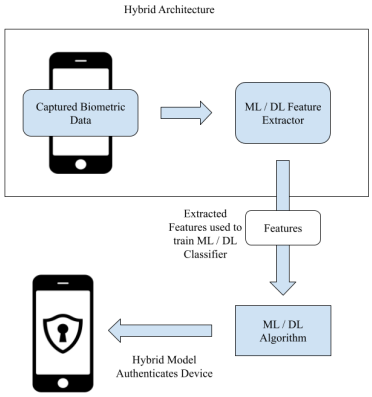

Figure 10. Depiction of a hybrid architecture.

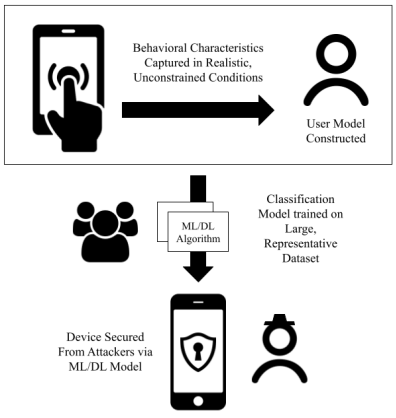

Figure 11. Depiction of the ideal behavioral biometric model.

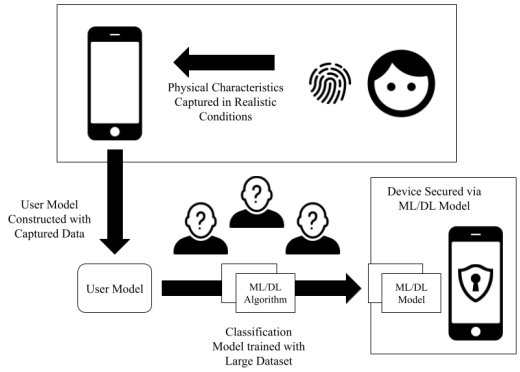

Figure 12. Depiction of the ideal physiological biometric model.

Author Contributions: Conceptualization, M.V. and R.D.; methodology, S.K.; validation, R.D., M.V.; formal analysis, R.D.; investigation, S.K.; resources, M.V.; data curation S.K.; writing—original draft preparation, S.K.; writing—review and editing, R.D. and M.V.; visualization, S.K.; supervision, R.D. and M.V.; project administration, R.D.; funding acquisition, M.V. All authors have read and agreed to the published version of the manuscript.

Funding: Funding for this project has been provided by the University of Wisconsin-Eau Claire's Blugold Fellowship and University of Wisconsin-Eau Claire Computer Science Department's Karlgaard Scholarship.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Silasai, O.; Khowfa, W. The Study on Using Biometric Authentication on Mobile Device. *NU Int. J. Sci.* **2020**, *17*, 90–110.
2. Siddiqui, N.; Pryor, L.; Dave, R. User authentication schemes using machine learning methods—A review. In Proceedings of the International Conference on Communication and Computational Technologies: ICCCT, Singapore, 28 August 2021. [CrossRef]
3. López, A.B. Deep learning in biometrics: A survey. *ADCAIJ Adv. Distrib. Comput. Artif. Intell. J.* **2019**, *8*, 19–32. [CrossRef]
4. Kokal, S.; Pryor, L.; Dave, R. Exploration of Machine Learning Classification Models Used for Behavioral Biometrics Authentication. In Proceedings of the 2022 8th International Conference on Computer Technology Applications, Vienna, Austria, 12 May 2022.
5. Dahia, G.; Jesus, L.; Pamplona, S.M. Continuous authentication using biometrics: An advanced review. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2020**, *10*, e1365. [CrossRef]
6. Jaseena, K.U.; Kovoor, B.C. A survey on deep learning techniques for big data in biometrics. *Int. J. Adv. Res. Comput. Sci.* **2018**, *9*, 12–17. [CrossRef]
7. Wang, H.; He, H.; Song, C.; Tang, H.; Sun, Y.; Qiao, Y.; Zhang, W. Who Is Using the Phone? Representation-Learning-Based Continuous Authentication on Smartphones. *Secur. Commun. Netw.* **2022**, *2022*, 6339407. [CrossRef]
8. Yang, W.; Wang, M.; Zou, S.; Peng, J.; Xu, G. An Implicit Identity Authentication Method Based on Deep Connected Attention CNN for Wild Environment. In Proceedings of the 9th International Conference on Communications and Broadband Networking, Shanghai, China, 25 February 2021. [CrossRef]
9. Song, Y.; Cai, Z. Integrating Handcrafted Features with Deep Representations for Smartphone Authentication. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2022**, *6*, 1–27. [CrossRef]
10. Acien, A.; Morales, A.; Vera-Rodriguez, R.; Fierrez, J. Smartphone sensors for modeling human-computer interaction: General outlook and research datasets for user authentication. In Proceedings of the 2020 IEEE 44th Annual Computers, Software, and Applications Conference, Madrid, Spain, 13 July 2020. [CrossRef]
11. Tolosana, R.; Vera-Rodriguez, R.; Fierrez, J.; Ortega-Garcia, J. BioTouchPass2: Touchscreen password biometrics using time-aligned recurrent neural networks. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 2616–2618. [CrossRef]
12. Volaka, H.C.; Alptekin, G.; Basar, O.E.; Isbilen, M.; Incel, O.D. Towards continuous authentication on mobile phones using deep learning models. *Procedia Comput. Sci.* **2019**, *155*, 177–184. [CrossRef]
13. Al-Dori, A.S. Touchscreen-based Smartphone Continuous Authentication System (SCAS) using Deep Neural Network. *TURCOMAT* **2021**, *12*, 2382–2391.
14. Liang, X.; Zou, F.; Li, L.; Yi, P. Mobile terminal identity authentication system based on behavioral characteristics. *Int. J. Distrib. Sens. Netw.* **2020**, *16*, 1550147719899371. [CrossRef]
15. Samet, S.; Ishraque, M.T.; Ghadamyari, M.; Kakadiya, K.; Mistry, Y.; Nakkabi, Y. TouchMetric: A machine learning based continuous authentication feature testing mobile application. *Int. J. Inf. Technol.* **2019**, *11*, 625–631. [CrossRef]
16. Chang, I.; Low, C.Y.; Choi, S.; Teoh, A.B. Kernel deep regression network for touch-stroke dynamics authentication. *IEEE Signal Process. Lett.* **2018**, *25*, 1109–1113. [CrossRef]
17. Nader, J.; Alsadoon, A.; Prasad, P.W.; Singh, A.K.; Elchouemi, A. Designing touch-based hybrid authentication method for smartphones. *Procedia Comput. Sci.* **2015**, *70*, 198–204. [CrossRef]
18. Deb, D.; Guirguis, M.M. Use of auxiliary classifier generative adversarial network in touchstroke authentication. In Proceedings of the 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 14 December 2020. [CrossRef]
19. Shankar, V.; Singh, K. An intelligent scheme for continuous authentication of smartphone using deep auto encoder and softmax regression model easy for user brain. *IEEE Access* **2019**, *7*, 48645–48654. [CrossRef]
20. Torres, J.; Santos, S.; Alepis, E.; Patsakis, C. Behavioral Biometric Authentication in Android Unlock Patterns through Machine Learning. In Proceedings of the 5th International Conference on Information Systems Security and Privacy (ICISSP), Prague, Czech Republic, 23–25 February 2019. [CrossRef]
21. Shi, D.; Tao, D.; Wang, J.; Yao, M.; Wang, Z.; Chen, H.; Helal, S. Fine-grained and context-aware behavioral biometrics for pattern lock on smartphones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2021**, *5*, 1–30. [CrossRef]

22. Fierrez, J.; Pozo, A.; Martinez-Diaz, M.; Galbally, J.; Morales, A. Benchmarking touchscreen biometrics for mobile authentication. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2720–2733. [CrossRef]
23. Li, W.; Meng, W.; Furnell, S. Exploring touch-based behavioral authentication on smartphone email applications in IoT-enabled smart cities. *Pattern Recognit. Lett.* **2021**, *144*, 35–41. [CrossRef]
24. Miyamoto, N.; Shibata, C.; Kinoshita, T. Authentication by Touch Operation on Smartphone with Support Vector Machine. *Int. J. Inf. Secur. Res* **2017**, *7*, 725–733. [CrossRef]
25. Antal, M.; Szabó, L.Z. Biometric authentication based on touchscreen swipe patterns. *Procedia Technol.* **2016**, *22*, 862–869. [CrossRef]
26. Buriro, A.; Gupta, S.; Yautsiukhin, A.; Crispo, B. Risk-driven behavioral biometric-based one-shot-cum-continuous user authentication scheme. *J. Signal Process. Syst.* **2021**, *93*, 989–1006. [CrossRef]
27. Zhang, X.; Zhang, P.; Hu, H. Multimodal continuous user authentication on mobile devices via interaction patterns. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 5677978. [CrossRef]
28. Leyfer, K.; Spivak, A. Continuous user authentication by the classification method based on the dynamic touchscreen biometrics. In Proceedings of the 2019 24th Conference of Open Innovations Association (FRUCT), Moscow, Russia, 8 April 2019. [CrossRef]
29. Levi, M.; Hazan, I.; Agmon, N.; Eden, S. Behavioral embedding for continuous user verification in global settings. *Comput. Secur.* **2022**, *119*, 102716. [CrossRef]
30. Lee, J.; Park, S.; Kim, Y.G.; Lee, E.K.; Jo, J. Advanced Authentication Method by Geometric Data Analysis Based on User Behavior and Biometrics for IoT Device with Touchscreen. *Electronics* **2021**, *10*, 2583. [CrossRef]
31. Huang, E.; Troia, F.D.; Stamp, M. Evaluating deep learning models and adversarial attacks on accelerometer-based gesture authentication. *Artif. Intell. Cybersecur.* **2022**, *54*, 243–259. [CrossRef]
32. Li, Y.; Hu, H.; Zhu, Z.; Zhou, G. SCANet: Sensor-based continuous authentication with two-stream convolutional neural networks. *ACM Trans. Sens. Netw. (TOSN)* **2020**, *16*, 1–27. [CrossRef]
33. Centeno, M.P.; Guan, Y.; van Moorsel, A. Mobile based continuous authentication using deep features. In Proceedings of the 2nd International Workshop on Embedded and Mobile Deep Learning, New York, NY, USA, 15 June 2018. [CrossRef]
34. Benegui, C.; Ionescu, R.T. Convolutional neural networks for user identification based on motion sensors represented as images. *IEEE Access* **2020**, *8*, 61255–61266. [CrossRef]
35. Zhu, T.; Weng, Z.; Chen, G.; Fu, L. A hybrid deep learning system for real-world mobile user authentication using motion sensors. *Sensors* **2020**, *20*, 3876. [CrossRef]
36. Li, Y.; Tao, P.; Deng, S.; Zhou, G. DeFFusion: CNN-based continuous authentication using deep feature fusion. *ACM Trans. Sens. Netw. (TOSN)* **2021**, *18*, 1–20. [CrossRef]
37. Abuhamad, M.; Abuhmed, T.; Mohaisen, D.; Nyang, D. AUToSen: Deep-learning-based implicit continuous authentication using smartphone sensors. *IEEE Internet Things J.* **2020**, *7*, 5008–5020. [CrossRef]
38. Mekruksavanich, S.; Jitpattanakul, A. Deep learning approaches for continuous authentication based on activity patterns using mobile sensing. *Sensors* **2021**, *21*, 7519. [CrossRef]
39. Amini, S.; Noroozi, V.; Pande, A.; Gupte, S.; Yu, P.S.; Kanich, C. Deepauth: A framework for continuous user re-authentication in mobile apps. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Torino, Italy, 17 October 2018. [CrossRef]
40. Neverova, N.; Wolf, C.; Lacey, G.; Fridman, L.; Chandra, D.; Barbello, B.; Taylor, G. Learning human identity from motion patterns. *IEEE Access* **2016**, *4*, 1810–1820. [CrossRef]
41. Centeno, M.P.; van Moorsel, A.; Castruccio, S. Smartphone continuous authentication using deep learning autoencoders. In Proceedings of the 2017 15th Annual Conference on Privacy, Security and Trust (pst), Calgary, AB, Canada, 28 August 2017. [CrossRef]
42. Li, C.; Jing, J.; Liu, Y. Mobile user authentication-Turn it to unlock. In Proceedings of the 2021 6th International Conference on Mathematics and Artificial Intelligence, Chengdu, China, 19 March 2021. [CrossRef]
43. Ehatisham-ul-Haq, M.; Azam, M.A.; Naeem, U.; Amin, Y.; Loo, J. Continuous authentication of smartphone users based on activity pattern recognition using passive mobile sensing. *J. Netw. Comput. Appl.* **2018**, *109*, 24–35. [CrossRef]
44. Zhu, T.; Qu, Z.; Xu, H.; Zhang, J.; Shao, Z.; Chen, Y.; Prabhakar, S.; Yang, J. RiskCog: Unobtrusive real-time user authentication on mobile devices in the wild. *IEEE Trans. Mob. Comput.* **2019**, *19*, 466–483. [CrossRef]
45. Hernández-Álvarez, L.; De Fuentes, J.M.; González-Manzano, L.; Encinas, L.H. SmartCAMPP-Smartphone-based continuous authentication leveraging motion sensors with privacy preservation. *Pattern Recognit. Lett.* **2021**, *147*, 189–196. [CrossRef]
46. Maghsoudi, J.; Tappert, C.C. Increasing Accuracy Rate of Behavioural Biometrics for User Authentication on Android-Based Smartphones. In Proceedings of the Student-Faculty Research Day, Pace University, New York, NY, USA, 5 May 2017.
47. Malik, M.N.; Azam, M.A.; Ehatisham-Ul-Haq, M.; Ejaz, W.; Khalid, A. ADLAuth: Passive authentication based on activity of daily living using heterogeneous sensing in smart cities. *Sensors* **2019**, *19*, 2466. [CrossRef]
48. Buriro, A.; Crispo, B.; Conti, M. AnswerAuth: A bimodal behavioral biometric-based user authentication scheme for smartphones. *J. Inf. Secur. Appl.* **2019**, *44*, 89–103. [CrossRef]
49. Stragapede, G.; Vera-Rodriguez, R.; Tolosana, R.; Morales, A.; Acien, A.; Le Lan, G. Mobile behavioral biometrics for passive authentication. *Pattern Recognit. Lett.* **2022**, *157*, 35–41. [CrossRef]

50. Deb, D.; Ross, A.; Jain, A.K.; Prakah-Asante, K.; Prasad, K.V. Actions speak louder than (pass) words: Passive authentication of smartphone* users via deep temporal features. In Proceedings of the 2019 International Conference on Biometrics (ICB), Crete, Greece, 4 June 2019. [CrossRef]
51. Tse, K.W.; Hung, K. Framework for user behavioural biometric identification using a multimodal scheme with keystroke trajectory feature and recurrent neural network on a mobile platform. *IET Biom.* **2022**, *11*, 157–170. [CrossRef]
52. Acien, A.; Morales, A.; Monaco, J.V.; Vera-Rodriguez, R.; Fierrez, J. TypeNet: Deep learning keystroke biometrics. *IEEE Trans. Biom. Behav. Identity Sci.* **2021**, *4*, 57–70. [CrossRef]
53. Sun, Y.; Gao, Q.; Du, X.; Gu, Z. Smartphone user authentication based on holding position and touch-typing biometrics. *Comput. Mater. Contin.* **2019**, *3*, 1365–1375. [CrossRef]
54. Stragapede, G.; Vera-Rodriguez, R.; Tolosana, R.; Morales, A. BehavePassDB: Benchmarking Mobile Behavioral Biometrics. *arXiv* **2022**, arXiv:2206.02502.
55. Stragapede, G.; Delgado-Santos, P.; Tolosana, R.; Vera-Rodriguez, R.; Guest, R.; Morales, A. Mobile Keystroke Biometrics Using Transformers. In Proceedings of the 2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG), Waikoloa Beach, HI, USA, 5–8 January 2023. [CrossRef]
56. Deng, Y.; Zhong, Y. Keystroke dynamics advances for mobile devices using deep neural network. *Recent Adv. User Authentication Using Keystroke Dyn. Biom.* **2015**, *2*, 59–70. [CrossRef]
57. Sun, L.; Cao, B.; Wang, J.; Srisa-an, W.; Philip, S.Y.; Leow, A.D.; Checkoway, S. Kollector: Detecting fraudulent activities on mobile devices using deep learning. *IEEE Trans. Mob. Comput.* **2020**, *20*, 1465–1476. [CrossRef]
58. Salem, A.; Zaidan, D.; Swidan, A.; Saifan, R. Analysis of strong password using keystroke dynamics authentication in touch screen devices. In Proceedings of the 2016 Cybersecurity and Cyberforensics Conference (CCC), Amman, Jordan, 2 August 2016. [CrossRef]
59. Alshanketi, F.; Traore, I.; Ahmed, A.A. Improving performance and usability in mobile keystroke dynamic biometric authentication. In Proceedings of the 2016 IEEE Security and Privacy Workshops (SPW), San Jose, CA, USA, 22 May 2016. [CrossRef]
60. Hriez, S.; Obeid, N.; Awajan, A. User authentication on smartphones using keystroke dynamics. In Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems, New York, NY, USA, 2 December 2019. [CrossRef]
61. Krishnamoorthy, S.; Rueda, L.; Saad, S.; Elmiligi, H. Identification of User Behavioral Biometrics for Authentication Using Keystroke Dynamics and Machine Learning. In Proceedings of the 2018 2nd International Conference on Biometric Engineering and Applications, Amsterdam, The Netherlands, 16 May 2018. [CrossRef]
62. Nerini, M.; Favarelli, E.; Chiani, M. Augmented PIN Authentication through Behavioral Biometrics. *Sensors* **2022**, *22*, 4857. [CrossRef]
63. Gadaleta, M.; Rossi, M. Idnet: Smartphone-based gait recognition with convolutional neural networks. *Pattern Recognit.* **2018**, *74*, 25–37. [CrossRef]
64. Wang, C.; Xiao, Y.; Gao, X.; Li, L.; Wang, J. A framework for behavioral biometric authentication using deep metric learning on mobile devices. *IEEE Trans. Mob. Comput.* **2021**, *22*, 19–36. [CrossRef]
65. Middya, A.I.; Roy, S.; Mandal, S.; Talukdar, R. Privacy protected user identification using deep learning for smartphone-based participatory sensing applications. *Neural Comput. Appl.* **2021**, *33*, 17303–17313. [CrossRef]
66. Watanabe, Y.; Kimura, M. Gait identification and authentication using LSTM based on 3-axis accelerations of smartphone. *Procedia Comput. Sci.* **2020**, *176*, 3873–3880. [CrossRef]
67. Hu, G.; He, Z.; Lee, R.B. Smartphone impostor detection with behavioral data privacy and minimalist hardware support. *arXiv* **2021**, arXiv:2103.06453.
68. Chauhan, J.; Kwon, Y.D.; Hui, P.; Mascolo, C. Contauth: Continual learning framework for behavioral-based user authentication. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2020**, *4*, 1–23. [CrossRef]
69. Zou, Q.; Wang, Y.; Wang, Q.; Zhao, Y.; Li, Q. Deep learning-based gait recognition using smartphones in the wild. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 3197–3212. [CrossRef]
70. Zeng, X.; Zhang, X.; Yang, S.; Shi, Z.; Chi, C. Gait-based implicit authentication using edge computing and deep learning for mobile devices. *Sensors* **2021**, *21*, 4592. [CrossRef] [PubMed]
71. Anh Khoa, T.; The Truong, D.N.; Dang, D.N. Cross-Modal Deep Neural Networks based Smartphone Authentication for Intelligent Things System. In Proceedings of the 2021 Workshop on Intelligent Cross-Data Analysis and Retrieval, New York, NY, USA, 21 August 2021. [CrossRef]
72. Delgado-Santos, P.; Tolosana, R.; Guest, R.; Vera-Rodriguez, R.; Deravi, F.; Morales, A. GaitPrivacyON: Privacy-preserving mobile gait biometrics using unsupervised learning. *Pattern Recognit. Lett.* **2022**, *161*, 30–37. [CrossRef]
73. Cao, Q.; Xu, F.; Li, H. User Authentication by Gait Data from Smartphone Sensors Using Hybrid Deep Learning Network. *Mathematics* **2022**, *10*, 2283. [CrossRef]
74. Zeroual, A.; Amroune, M.; Derdour, M.; Meraoumia, A.; Bentahar, A. Deep authentication model in Mobile Cloud Computing. In Proceedings of the 2018 3rd International Conference on Pattern Analysis and Intelligent Systems (PAIS), Tebessa, Algeria, 24 October 2018. [CrossRef]
75. Azouji, N.; Sami, A.; Taheri, M. EfficientMask-Net for face authentication in the era of COVID-19 pandemic. *Signal Image Video Process.* **2022**, *16*, 1991–1999. [CrossRef]

76. Chen, S.; Liu, Y.; Gao, X.; Han, Z. Mobilefacenets: Efficient CNNs for Accurate Real-Time Face Verification on Mobile Devices. In Proceedings of the Chinese Conference on Biometric Recognition (CCBR), Urumqi, China, 11 August 2018. [CrossRef]
77. Oza, P.; Patel, V.M. Active authentication using an autoencoder regularized cnn-based one-class classifier. In Proceedings of the 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), Lille, France, 14 May 2019. [CrossRef]
78. Samangouei, P.; Chellappa, R. Convolutional neural networks for attribute-based active authentication on mobile devices. In Proceedings of the 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), Niagara Falls, NY, USA, 6 September 2016. [CrossRef]
79. Zhou, B.; Lohokare, J.; Gao, R.; Ye, F. EchoPrint: Two-Factor Authentication Using Acoustics and Vision on Smartphones. In Proceedings of the 24th Annual International Conference on Mobile Computing and Networking, New Delhi, India, 15 October 2018. [CrossRef]
80. Ríos-Sánchez, B.; Silva, D.C.; Martín-Yuste, N.; Sánchez-Ávila, C. Deep learning for face recognition on mobile devices. *IET Biom.* **2020**, *9*, 109–117. [CrossRef]
81. Saied, M.; Elshenawy, A.; Ezz, M.M. A Novel Approach for Improving Dynamic Biometric Authentication and Verification of Human Using Eye Blinking Movement. *Wirel. Pers. Commun.* **2020**, *115*, 859–876. [CrossRef]
82. Zeroual, A.; Amroune, M.; Derdour, M.; Bentahar, A. Lightweight deep learning model to secure authentication in Mobile Cloud Computing. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 6938–6948. [CrossRef]
83. Dar, S.A.; Palanivel, S. Real-Time Face Authentication Using Denoised Autoencoder (DAE) for Mobile Devices. In *Handbook of Research on Evolving Designs and Innovation in ICT and Intelligent Systems for Real-World Applications*; IGI Global: Hershey, PA, USA, 2022; pp. 163–176. [CrossRef]
84. Gunasinghe, H.; Bertino, E. PrivBioMTAuth: Privacy Preserving Biometrics-based and User Centric Protocol for User Authentication From Mobile Phones. *IEEE Trans. Inf. Forensics Secur.* **2017**, *13*, 1042–1057. [CrossRef]
85. Samangouei, P.; Patel, V.M.; Chellappa, R. Facial Attributes for Active Authentication on Mobile Devices. *Image Vis. Comput.* **2017**, *58*, 181–192. [CrossRef]
86. Garg, R.; Baweja, Y.; Ghosh, S.; Singh, R.; Vatsa, M.; Ratha, N. Heterogeneity aware deep embedding for mobile periocular recognition. In Proceedings of the 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS), Redondo Beach, CA, USA, 22 October 2018. [CrossRef]
87. Reddy, N.; Rattani, A.; Derakhshani, R. Ocularnet: Deep patch-based ocular biometric recognition. In Proceedings of the 2018 IEEE International Symposium on Technologies for Homeland Security (HST), Woburn, MA, USA, 23 October 2018. [CrossRef]
88. Nguyen, H.; Reddy, N.; Rattani, A.; Derakhshani, R. VISOB 2.0-the second international competition on mobile ocular biometric recognition. In Proceedings of the ICPR International Workshops and Challenges, Virtual Event, 10–15 January 2021. [CrossRef]
89. Reddy, N.; Rattani, A.; Derakhshani, R. Comparison of deep learning models for biometric-based mobile user authentication. In Proceedings of the 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS), Redondo Beach, CA, USA, 22 October 2018. [CrossRef]
90. Sandhya, M.; Morampudi, M.K.; Pruthweraaj, I.; Garepally, P.S. Multi-instance cancelable iris authentication system using triplet loss for deep learning models. *Vis. Comput.* **2022**, *39*, 1571–1581. [CrossRef]
91. Zheng, S.; Rahmat, R.W.; Khalid, F.; Nasharuddin, N.A. Learning scale-variant features for robust iris authentication with deep learning based ensemble framework. *arXiv* **2019**, arXiv:1912.00756. Available online: https://doi.org/10.48550/arXiv.1912.00756 (accessed on 11 April 2023).
92. Zeng, F.; Hu, S.; Xiao, K. Research on partial fingerprint recognition algorithm based on deep learning. *Neural Comput. Appl.* **2019**, *31*, 4789–4798. [CrossRef]
93. Wu, C.; He, K.; Chen, J.; Zhao, Z.; Du, R. Liveness is Not Enough: Enhancing Fingerprint Authentication with Behavioral Biometrics to Defeat Puppet Attacks. In Proceedings of the USENIX Security Symposium, Boston, MA, USA, 12 August 2020.
94. Arora, S.; Bhatia, M.S. Fingerprint spoofing detection to improve customer security in mobile financial applications using deep learning. *Arab. J. Sci. Eng.* **2020**, *45*, 2847–2863. [CrossRef]
95. Grosz, S.A.; Engelsma, J.J.; Liu, E.; Jain, A.K. C2cl: Contact to contactless fingerprint matching. *IEEE Trans. Inf. Forensics Secur.* **2021**, *17*, 196–210. [CrossRef]
96. Garcia-Martin, R.; Sanchez-Reillo, R. Deep learning for vein biometric recognition on a smartphone. *IEEE Access* **2021**, *9*, 98812–98832. [CrossRef]
97. Thullier, F.; Bouchard, B.; Menelas, B.A. A text-independent speaker authentication system for mobile devices. *Cryptography* **2017**, *1*, 16. [CrossRef]
98. Chen, Y.; Xue, M.; Zhang, J.; Guan, Q.; Wang, Z.; Zhang, Q.; Wang, W. Chestlive: Fortifying voice-based authentication with chest motion biometric on smart devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2021**, *5*, 1–25. [CrossRef]