

# Fingerprints as Predictors of Schizophrenia: A Deep Learning Study

Raymond Salvador<sup>\*,1,2,18,Ⓞ</sup>, María Ángeles García-León<sup>1,2,18</sup>, Isabel Feria-Raposo<sup>1,3,4</sup>, Carlota Botillo-Martín<sup>5</sup>, Carlos Martín-Lorenzo<sup>5</sup>, Carmen Corte-Souto<sup>6</sup>, Tania Aguilar-Valero<sup>6</sup>, David Gil-Sanz<sup>7,8,Ⓞ</sup>, David Porta-Pelayo<sup>9</sup>, Manuel Martín-Carrasco<sup>10</sup>, Francisco del Olmo-Romero<sup>11,12</sup>, Jose Maria Santiago-Bautista<sup>11</sup>, Pilar Herrero-Muñecas<sup>13</sup>, Eglee Castillo-Oramas<sup>14</sup>, Jesús Larrubia-Romero<sup>15</sup>, Zoila Rios-Alvarado<sup>3</sup>, José Antonio Larráz-Romeo<sup>16</sup>, María Guardiola-Ripoll<sup>1,2</sup>, Carmen Almodóvar-Payá<sup>1,2</sup>, Mar Fatjó-Vilas Mestre<sup>1,2,17</sup>, Salvador Sarró<sup>1,2</sup>, Peter J. McKenna<sup>1,2</sup>, HHFingerprints Group<sup>19</sup>, and Edith Pomarol-Clotet<sup>1,2</sup>

<sup>1</sup>FIDMAG Germanes Hospitalàries Research Foundation, Barcelona, Spain; <sup>2</sup>Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain; <sup>3</sup>Benito Menni Complex Assistencial en Salut Mental, Barcelona, Spain; <sup>4</sup>Unidad de Investigación en Cuidados y Servicios de Salud, Instituto de Salud Carlos III (Investén-ISCIII), Madrid, Spain; <sup>5</sup>Centro Sociosanitario Hermanas Hospitalarias, Palencia, Spain; <sup>6</sup>Hospital Mare de Déu de la Mercè, Barcelona, Spain; <sup>7</sup>Centro Hospitalario Padre Menni, Santander, Spain; <sup>8</sup>Universidad Europea del Atlántico, Santander, Spain; <sup>9</sup>Complejo asistencial Hermanas Hospitalarias, Málaga, Spain; <sup>10</sup>Psychiatric Clinic Padre Menni, Pamplona, Navarra, Spain; <sup>11</sup>Complejo Asistencial Benito Menni, Ciempozuelos, Madrid, Spain; <sup>12</sup>Clínica San Miguel Hermanas Hospitalarias, Madrid, Spain; <sup>13</sup>Hospital Sagrat Cor, Martorell, Barcelona, Spain; <sup>14</sup>Hospital Aita Menni, Mondragón, País Vasco; <sup>15</sup>Centro Neuropsiquiátrico Nuestra Señora del Carmen, Zaragoza, Spain; <sup>16</sup>Hermanas Hospitalarias, Provincia de España, Spain; <sup>17</sup>Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain

\*To whom correspondence should be addressed; FIDMAG Germanes Hospitalàries Research Foundation, C/Dr. Antoni Pujadas, 38, 08830, Sant Boi de Llobregat, Barcelona, Spain, e-mail: [rsalvador@fidmag.com](mailto:rsalvador@fidmag.com)

<sup>18</sup>These authors contributed equally to this work.

<sup>19</sup>**HHFingerprints Group:** Emilio González-Pablos, Emilio Negro-González, Eva María Castells Bescos, Elena Felipe Martínez, Paula Muñoz Hermoso, Cora Camaño Serna, Carlos Rebolleda Gil, Carmen Feliz Muñoz, Paula Sevillano De La Fuente, Manuel Sánchez Perez, Izascun Arrece Iriondo, José Vicente Jauregui Berecibar, Ana Domínguez Panchón, Alfredo Felices de la Fuente, Clara Bosque Gabarre.

**Background and Hypothesis:** The existing developmental bond between fingerprint generation and growth of the central nervous system points to a potential use of fingerprints as risk markers in schizophrenia. However, the high complexity of fingerprints geometrical patterns may require flexible algorithms capable of characterizing such complexity. **Study Design:** Based on an initial sample of scanned fingerprints from 612 patients with a diagnosis of non-affective psychosis and 844 healthy subjects, we have built deep learning classification algorithms based on convolutional neural networks. Previously, the general architecture of the network was chosen from exploratory fittings carried out with an independent fingerprint dataset from the National Institute of Standards and Technology. The network architecture was then applied for building classification algorithms (patients vs controls) based on single fingers and multi-input models. Unbiased estimates of classification accuracy were obtained by applying a 5-fold cross-validation scheme. **Study Results:** The highest level of accuracy from networks based on single fingers was

achieved by the right thumb network (weighted validation accuracy = 68%), while the highest accuracy from the multi-input models was attained by the model that simultaneously used images from the left thumb, index and middle fingers (weighted validation accuracy = 70%). **Conclusion:** Although fitted models were based on data from patients with a well established diagnosis, since fingerprints remain lifelong stable after birth, our results imply that fingerprints may be applied as early predictors of psychosis. Specially, if they are used in high prevalence subpopulations such as those of individuals at high risk for psychosis.

**Key words:** schizophrenia/machine learning/dermatoglyphics/diagnosis/artificial intelligence

## Introduction

Prenatal alterations in the development of the central nervous system of genetic and environmental origin have been suggested as possible causes for schizophrenia.<sup>1,2</sup>

Due to a common ontogenetic origin in the ectoderm (ie, the embryonic tissue that later differentiates to form epithelial and neural tissues) such alterations may also be reflected as abnormalities in finger skin patterns (also known as dermatoglyphs).<sup>3,4</sup> Within weeks 6 to 24 of gestation dermatoglyphs develop in parallel with neural migration processes.<sup>5</sup> While dermatoglyphic development is not fully understood, there is clear evidence of its heritability<sup>6-8</sup> as well as the influence of the intrauterine environment.<sup>9,10</sup> Factors such as chromosomal alterations, viral infections, or maternal stress during pregnancy have been associated with simplified and abnormal dermatoglyphic patterns<sup>8,11</sup> while the speed of fetal development has been also related to dermatoglyphic development.<sup>12</sup>

After gestation, fingerprints become lifelong stable and may be considered as indirect markers of alterations in early neurodevelopment. This has led to a significant amount of research looking for dermatoglyphic alterations in patients with schizophrenia, summarized in the meta-analysis of Golembo-Smith et al.<sup>4</sup> Studies included have frequently reported reductions in the number of dermopapillary ridges and higher levels of fluctuating asymmetry.<sup>13-15</sup> However, in most of them sample sizes were moderate or small, observed effect sizes were low and negative or conflicting results were present.<sup>4</sup>

To some extent, the uncertainty in these findings may also be due to the simplicity of the measured features, which probably were poor descriptors of the complex patterns present in dermatoglyphs. This limitation, though, may be overcome by using new tools recently developed in the field of machine learning and image processing such as the Deep Learning (DL) neural networks.<sup>16</sup> The architecture of DL networks allows modeling any type of mathematical relationship and are the ideal tool for characterizing patterns of high complexity “hidden” in the data.<sup>17</sup> In the field of medical imaging DL networks have shown their effectivity in a wide range of areas including detection of structures and organs, tissue segmentation, automatic detection of abnormalities and computer-aided diagnostics (see examples in reviews by Shen et al.,<sup>18</sup> Litjens et al.,<sup>19</sup> and Decuyper et al.<sup>20</sup>). In most image-based applications a specific type of DL network, known as convolutional neural network (CNN), has been used. CNNs, unlike other networks, explicitly take spatial contextual information into account.<sup>16,19</sup>

In this study, we fit CNNs to fingerprint images from a large sample including subjects with a diagnosis of non-affective psychosis and healthy individuals. In a preliminary step, the hyperparameters specifying the architecture of the network were selected by fitting CNNs with an independent sample of fingerprints from the National Institute of Standards and Technology (NIST). Once the network architecture was established, models based on single fingers were built to classify our sample of patients and controls, providing information

on the predictive power of each individual finger. Next, to maximize prediction accuracies multi-input models combining the information from several fingers were also fit. In all models a 5-fold cross-validation scheme was used to obtain non-biased estimates of diagnosis accuracy.

## Methods

### Sample

An initial sample of 612 patients with a diagnosis of schizophrenia ( $N = 544$ ) and of schizoaffective disorder ( $N = 68$ ) according to DSM-IV criteria were recruited from 13 facilities belonging to the Hermanas Hospitalarias cluster of mental health hospitals in Spain (located in Barcelona, Palencia, Santander, Málaga, Madrid, Navarra, Guipúzcoa, and Zaragoza). A second sample of  $N = 844$  healthy controls was also recruited from non-medical hospital staff, their relatives and acquaintances, plus independent sources in the community, discarding potential participants with a first degree relative with a diagnosis of psychosis. In both samples, subjects were aged 18 years or above. Individuals not belonging to the European-Caucasian ethnicity but which had been originally included in the study ( $N = 77$ ) were later discarded. All subjects gave written informed consent and the study procedures were approved by the Comité de Ética de la Investigación de FIDMAG Hermanas Hospitalarias and adhered to the Declaration of Helsinki.

### Acquisition and Preprocessing of Fingerprints

Images from all fingers of participants were acquired with a Kojak PL fingerprint scanner (<https://integratedbiometrics.com/>) by a group of well trained people under the supervision of a single researcher. These images were saved in digital format and sent to the institution where image processing was performed. Then, a set of sequential steps were applied for quality check and improvement of images:

1. A Gabor filter was used iteratively (3 iterations) to minimize discontinuities in fingerprint images produced by skin scars and wounds (see an example in [Appendix 1](#)). Specifically, the method proposed by Hong et al.<sup>21</sup> and implemented in the python library (<https://pypi.org/project/fingerprint-enhancer>) was applied.
2. A visual inspection on the filtered images of all individuals was carried out, discarding images from individual fingers where scars and abnormalities were still noticeable after filtering. This step was performed by 2 researchers, being blind for both site and group. Specifically, the first researcher went through all images and labeled them as ok or bad/dubious. Next, the second researcher inspected the dubious images, which

**Table 1.** Absolute and Relative Amount of Filtered Images Discarded After Visual Inspection in Patients (Initial  $N = 612$ ) and in Healthy Controls (Initial  $N = 844$ ). Names of Fingers are Provided With Their Abbreviations as Used Throughout the Text

Finger (abbreviation)	Patients	Controls
Right thumb (R1)	230 (39 %)	90 (11 %)
Right index (R2)	200 (34 %)	66 (8 %)
Right middle (R3)	223 (38 %)	92 (12 %)
Right ring (R4)	263 (45 %)	119 (15 %)
Right little (R5)	310 (53 %)	148 (19 %)
Left thumb (L1)	242 (41 %)	114 (15 %)
Left index (L2)	212 (36 %)	59 (8 %)
Left middle (L3)	252 (43 %)	81 (10 %)
Left ring (L4)	301 (51 %)	144 (18 %)
Left little (L5)	325 (55 %)	178 (23 %)

in the majority of cases were eventually rejected. The relative amount of discarded images was substantial and is reported in [table 1](#).

- Images containing fingers extending beyond the first (distal interphalangeal) crease were cut to discard the extended portion.
- Finally, as required by DL algorithms, the margins of images were enlarged to have equal sizes. For that, the X and Y dimensions of the largest image of each finger were considered.

#### *Selection of Hyperparameters Defining the Network Architecture and the Learning Process*

The specific architecture of DL convolutional networks depends on a large number of parameters (known as hyperparameters) which have to be chosen before fitting the predictive models with actual data. These include, among others, the number of convolutional and dense layers, the size and number of filters in each layer, the inclusion of specific pooling, padding or strides between layers and the selection of the non-linear activation functions for intermediate and final layers. Also, additional decisions should be taken on the features ruling the training process, such as the loss function, the optimization algorithm, the usage and intensity of regularization or drop out, the number of epochs, and the size and number of batches.

The initial selection of these hyperparameters usually involves many attempts until an optimal configuration is found. Consequently, if the data used for this initial step is also used in the final network fittings, the risk of overfitting and of reporting biased (overoptimistic) classification accuracies becomes very high.<sup>16</sup> To avoid this, we used a publicly available set of fingerprint images from the National Institute of Standards and Technology (NIST) (<https://www.nist.gov/>). Specifically, 442 images from the right index were extensively used to find an optimal combination of hyperparameters for the prediction

of sex (which was the only dichotomic variable available in the dataset). Here it was assumed that a network architecture optimized for the classification of sex would also capture finger traits relevant for classifying patients and controls. Finally, to check if the selected hyperparameters were also suitable for our data, we carried out fittings using right index images from our healthy controls to classify them according to sex.

#### *Fitting of Deep Learning Networks for Patient/Control Classification*

Once all hyperparameters of the network were well established, 2 types of models were built for the patient/control classification using the scanned fingerprints ([figure 1](#)). On the one hand, predictive models based on single-finger images were fitted. For that, fingers with a rejection rate lower than 50% in patients were considered (see [table 1](#)). Specifically, individual models for fingers R1, R2, R3, R4, L1, L2, and L3 (see acronyms in [table 1](#)) were generated. On the other hand, networks combining information from several fingers were also considered. Following previous evidence pointing to a differential degree of right-left asymmetry in patients<sup>4</sup> models containing both right and left instances of the same finger were fitted. For that, two-finger models with right-left pairs R1-L1, R2-L2, and R3-L3 were considered. Three finger models including the thumb, index and middle finger of each hand (R1-R2-R3 and L1-L2-L3) were also trained, but more complex models or models including instances of the ring and little fingers were discarded as sample sizes became too small for reliable network development (fitting multi-finger models required having correct images from all fingers of each individual).

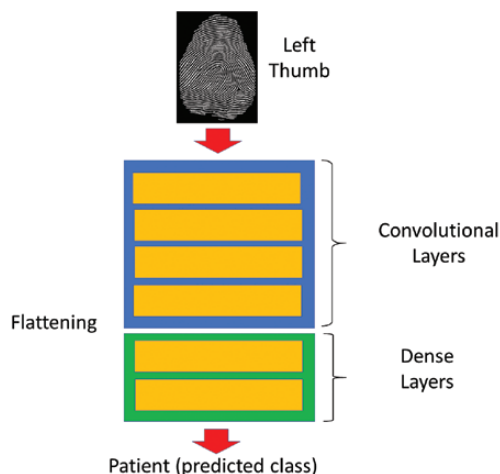
The same network architecture used in the single-finger models was considered as the building blocks for the multi-finger models ([figure 1](#)). Specifically, multi-input models, where each finger is processed by a parallel convolutional branch whose output is merged into a single dense layer containing the outputs of the other branches (fingers) were applied.<sup>22</sup>

From the initial fittings using the NIST dataset it became clear that sex had a strong influence on fingerprints. Hence, to avoid the confounding effect of sex on the patient/control classifications, samples of images had to be matched for sex (ie, they had to have the same proportion of males and females in patients and controls). Final sample sizes and sex ratios used for each model are provided in [table 2](#), together with information on age and estimates of premorbid IQ obtained through a Spanish word accentuation test based on thirty words.<sup>23</sup>

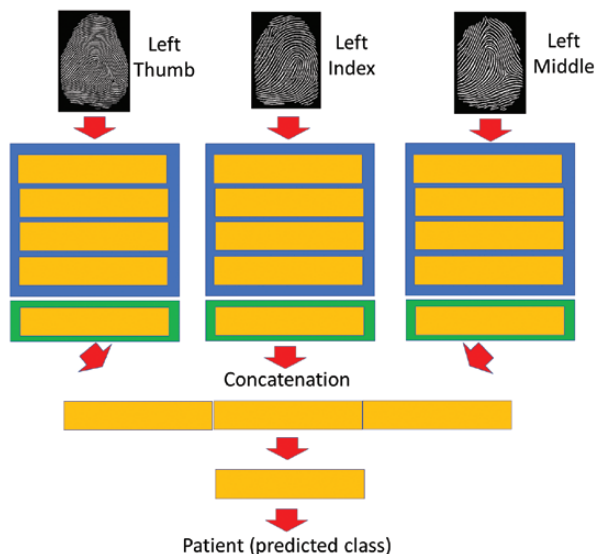
To obtain unbiased estimates of classification accuracies from the models, a 5-fold cross-validation scheme was applied. Final accuracy estimates were provided by averaged classification success values from validation subsamples obtained at the end of the last epoch.



## A) SINGLE FINGER CNN



## B) MULTI-INPUT CNN



**Fig. 1.** Two types of convolutional neural network (CNN) models were fitted with the fingerprints of patients and healthy controls. On the one hand, predictive models based on single-finger images were built (A). These models had a standard CNN architecture composed of a set of convolutional layers (ie, made of spatial convolution filters) followed by dense layers obtained by flattening (rearranging the convolutional structure to a simple 1-dimensional vector of fully connected nodes). The creation of these single-finger models followed a 2 step procedure. First, the general structure and main properties of the network (also known as hyperparameters) were chosen through trial and error on an independent external fingerprint dataset. Once the CNN architecture had been set, samples of finger images from patients and healthy controls obtained in this study were used to estimate the weights (coefficients) of the final predictive networks. (B) Considering several instances of this architecture (one instance for each finger) multi-input models that use images from several fingers were also built. At a certain level of the network, after the flattening step, the information from the different fingers had to be merged through concatenation of their dense layers into a single vector. In both single and multi-input models, a sigmoid

Bootstrap 95% intervals were derived from this same data. Building and training of all networks was carried out using the R API of Keras (<https://keras.rstudio.com>) which, in turn, runs on the TensorFlow platform ([www.tensorflow.org](http://www.tensorflow.org)). All network calculations were performed on a NVIDIA Quadro RTX 6000 24GB Graphics Processing Unit.

## Results

### *Hyperparameter Selection and Checking*

The general architecture of the network selected after exploratory fittings based on the NIST dataset, as listed by the Keras software, is shown in [Appendix 2](#). The chosen model contained 4 initial convolutional layers (three of them followed by pooling of the maximum value) and two dense layers. In all layers Rectified linear Units (relu) were used as activation functions, apart from the last dense layer where a sigmoid was considered to obtain outputs in the 0–1 range (ie, the range of probabilities). The number of convolution filters in each layer is also provided in [Appendix 2](#). Other relevant hyperparameters and settings are: optimization function = Root Mean Squared Propagation, loss function = Binary Cross-entropy, metrics = Accuracy, Number of epochs = 300, Batch Size = 32, convolution filter size =  $3 \times 3$ , pooling window size =  $4 \times 4$  and Drop Out regularization with a 0.5 rate.

Levels of accuracy in the validation subsets of each one of the 5 cross-validation folds, achieved by the network fittings as epochs were run, are shown in [Appendix 3A](#). As the NIST sample was not matched by sex (19% females vs 81% males) balanced (weighted) accuracies were used and reported instead of raw accuracies. Network fitting with the previously listed hyperparameter settings led to a weighted validation accuracy of 0.733 (0.681–0.787 bootstrap 95% interval).

When the same configuration of hyperparameters was used to classify sex in our subset of healthy subjects ( $N = 719$  correct images of right index in the whole sample; 35% men, 65% women) a lower but clearly significant weighted validation accuracy of 0.718 (0.684–0.752 bootstrap 95% interval) was achieved. [Appendix 3B](#) shows the evolution of weighted accuracies through the 300 epochs for the 5 validation subsets derived from the 5 cross-validation folds.

### *Diagnostic Classification Based on Single Fingers*

[Table 2](#) reports weighted validation accuracies achieved by fitting the previously selected network on each one of

---

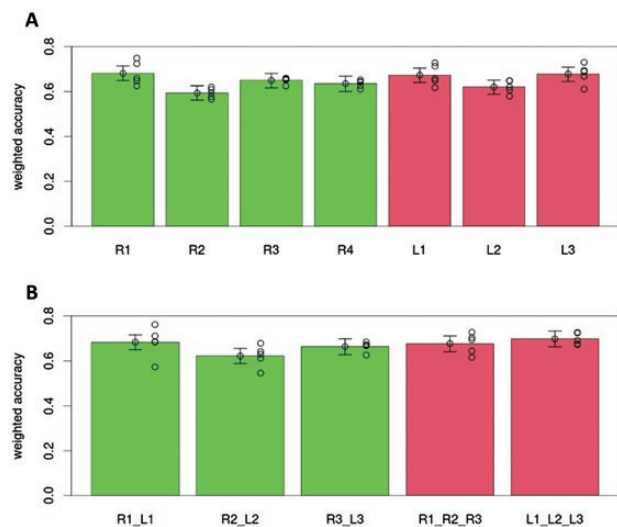
activation function in the last dense layer provided estimates of the probability for a test individual of belonging to the target population (ie, of being a patient).

**Table 2.** Final Sample Sizes, Age, Premorbid IQ as Provided by the Spanish Word Accentuation Test, Weighted Validation Accuracies, Sensitivities, Specificities and Area Under the Curve Values are Reported for Each Model. See Meaning of Finger Abbreviations in [table 1](#)

Fingers in Model	R1	R2	R3	R4	L1	L2
Sample sizes controls, patients	485, 359	507, 389	485, 366	492, 326	479, 347	490, 377
Sex ratios* male, female	50%, 50%	50%, 50%	50%, 50%	50%, 50%	50%, 50%	50%, 50%
Age in years**	52.56 (13.31)	52.52 (12.81)	51.87 (12.90)	51.46 (13.15)	51.50 (12.60)	51.88 (12.41)
Mean (SD)	44.59 (10.71)	44.51 (10.81)	44.73 (10.77)	44.24 (10.70)	44.39 (10.73)	44.78 (10.87)
Premorbid IQ**	17.71 (6.73)	17.76 (6.54)	17.91 (6.48)	18.03 (6.69)	17.71 (6.64)	17.98 (6.51)
Mean (SD)	23.88 (3.78)	23.82 (3.83)	23.93 (3.75)	23.81 (3.88)	23.80 (3.83)	23.94 (3.76)
Weighted accuracy	0.681	0.593	0.649	0.635	0.673	0.6205
(95% bootstrap CI)	(0.649-0.713)	(0.563-0.623)	(0.617-0.682)	(0.602-0.668)	(0.64-0.704)	(0.588-0.653)
Sensitivity	0.669	0.563	0.628	0.601	0.617	0.549
(95% bootstrap CI)	(0.624-0.714)	(0.517-0.611)	(0.578-0.68)	(0.549-0.655)	(0.569-0.669)	(0.499-0.601)
Specificity	0.693	0.623	0.67	0.669	0.729	0.692
(95% bootstrap CI)	(0.652-0.733)	(0.582-0.667)	(0.63-0.715)	(0.623-0.71)	(0.686-0.771)	(0.651-0.732)
Area under the curve	0.742	0.637	0.711	0.663	0.74	0.677
(95% bootstrap CI)	(0.707-0.775)	(0.602-0.672)	(0.678-0.746)	(0.626-0.7)	(0.706-0.774)	(0.64-0.712)
Fingers in Model	L3	R1_L1	R2_L2	R3_L3	R1_R2_R3	L1_L2_L3
Sample sizes controls, patients	507, 337	439, 286	465, 318	457, 296	435, 256	418, 262
Sex ratios* male, female	50%, 50%	50%, 50%	52%, 48%	52%, 48%	51%, 49%	53%, 47%
Age in years**	50.91 (12.43)	50.72 (12.81)	50.95 (12.04)	50.01 (12.43)	50.22 (12.37)	49.40 (11.96)
Mean (SD)	44.05 (10.60)	44.49 (10.74)	44.54 (10.87)	44.34 (10.75)	44.38 (10.67)	44.40 (10.78)
Premorbid IQ**	17.97 (6.53)	17.76 (6.69)	18.14 (6.43)	18.10 (6.45)	18.19 (6.46)	18.23 (6.54)
Mean (SD)	23.80 (3.81)	23.86 (3.78)	23.89 (3.78)	23.92 (3.75)	23.87 (3.74)	23.90 (3.81)
Weighted accuracy	0.678	0.6825	0.6215	0.6635	0.6765	0.6975
(95% bootstrap CI)	(0.647-0.709)	(0.649-0.717)	(0.586-0.654)	(0.629-0.697)	(0.641-0.713)	(0.664-0.733)
Sensitivity	0.638	0.643	0.563	0.625	0.613	0.668
(95% bootstrap CI)	(0.586-0.686)	(0.589-0.699)	(0.506-0.617)	(0.567-0.679)	(0.552-0.67)	(0.611-0.725)
Specificity	0.718	0.722	0.68	0.702	0.74	0.727
(95% bootstrap CI)	(0.681-0.756)	(0.681-0.763)	(0.638-0.721)	(0.662-0.744)	(0.699-0.784)	(0.685-0.77)
Area under the curve	0.734	0.748	0.673	0.732	0.74	0.769
(95% bootstrap CI)	(0.7-0.764)	(0.71-0.784)	(0.634-0.708)	(0.697-0.766)	(0.702-0.775)	(0.731-0.804)

\*Sex ratios are the same for patients and controls.

\*\*In all samples, mean differences in age and premorbid IQ between patients and controls as tested with *t*-tests were highly significant ( $P < .001$ ).



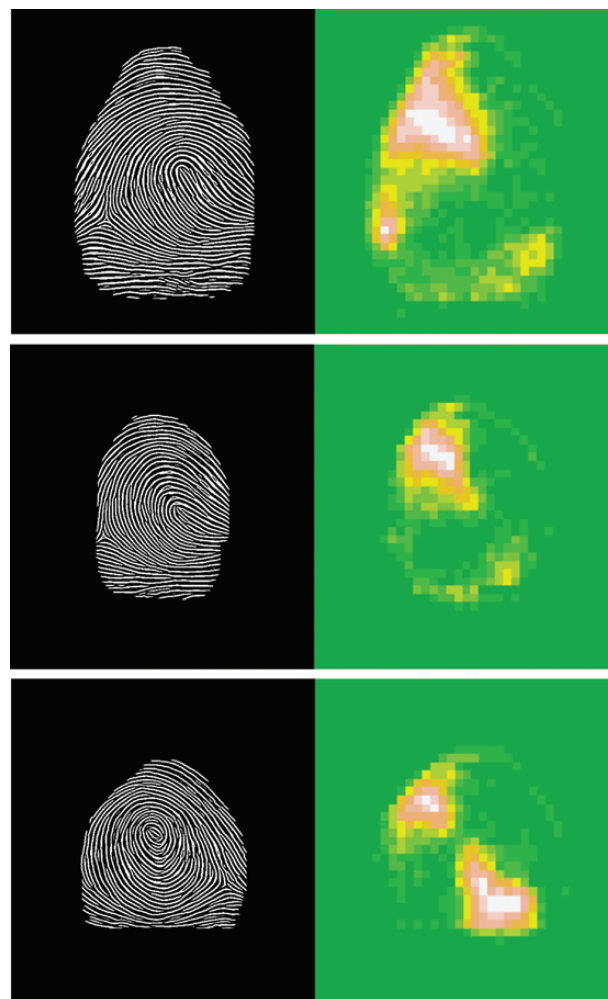
**Fig. 2.** (A) Upper plot shows the mean weighted accuracies achieved by each one of the networks based on a single fingerprint (with error bars based on bootstrap 95% confidence intervals). The plot also shows the 5 individual weighted accuracies estimated from each 1 of the 5 cross-validation test sets. (B) Analogous plot but for networks based on several fingers (multi-input models). Fingerprints are coded following the acronyms of [table 1](#).

the individual fingers separately (R1, R2, R3, R4, L1, L2, and L3). The highest accuracy (of 68%) was attained by the Right Thumb images, followed by the Left Middle and Left Thumb images (68% and 67%) while the lowest accuracy (of 59%) was provided by the Right Index. The remaining fingers had accuracies between these extreme values (see [figure 2A](#)).

To gain some insights on the inner workings of the developed neural networks, heatmaps of intermediate convolutional layers were created. Specifically, by calculating a weighted average of channels generated by a convolutional layer of the fitted model, maps of the areas of an individual fingerprint image that have been more relevant in the classification of a given subject may be derived (see [figure 3](#)).<sup>24</sup> Examples of heatmaps from correctly classified patients derived from each one of the fingers finally used in this study are shown in [Appendix 4](#). In general, regions with a substantial weight on the classification of individuals are quite large but are well delimited spatially. Although, for a given finger, there are noticeable interindividual similarities in the anatomical location of these influential areas, variability between individuals is also evident.

#### *Diagnostic Classification Based on Multi-Input Networks*

Finally, models based on more than one finger always provided weighted accuracies equaling or surpassing the maximum accuracy achieved by any of the individual fingers included in their respective models (see



**Fig. 3.** Left column: Examples of Right Thumb fingerprints for 3 patients correctly classified by the network. Right column: Heatmaps for these fingerprints obtained by a weighted average of channels generated by one of the intermediate convolutional layers of the predictive network (conv2d\_2). In these, the spatial distribution of anatomical areas with a greater weight in the classification of the 3 subjects is shown in light colors.

[table 2](#) and [figure 2B](#)). Specifically, the left-hand combination L1-L2-L3 led to the highest accuracy of all (70%).

#### **Discussion**

Our results suggest that fingerprints are valuable sources for diagnosis of non-affective psychosis and that CNNs are a feasible tool to achieve this goal. Since dermatoglyphic patterns, once formed, are stable through life it may also be inferred that fingerprint images are potentially useful for early prediction of risk before the disorder is developed. Although a maximum accuracy of 70% does not provide enough precision for a faultless diagnosis, fingerprint images may still be valuable, especially if they are combined with other sources of information that have already shown some predictive power in

schizophrenia such as genetics<sup>25,26</sup> and brain imaging data.<sup>27,28</sup> Indeed, the same multi-input approach used here to combine different fingers may be easily extended to include data sources of different natures.<sup>22</sup>

On the other hand, the 70% limit in accuracy also reflects the biological origin of fingerprint abnormalities. As stated in the introduction, dermatoglyphic development is restricted to a period during gestation and it can only be affected by genetic or environmental factors that were active during this period. While there is strong evidence for the effect of such factors, there are many other processes occurring later in life that have also been shown to increase the risk for schizophrenia<sup>29</sup> and which would expose a fingerprint based accuracy much higher than 70% as unrealistic.

Although a considerable effort was put on the proper design and execution of the study, some sources of potential bias have remained. Thus, as shown in [table 2](#), samples matched for sex ratio had substantial differences in pre-morbid IQ between patients and controls. While a major frequency or intensity in early neurodevelopmental events is likely to have led to reduced pre-morbid IQ values, we cannot completely rule out other sources of bias underlying such differences. The mean age in the matched (by sex ratio) samples was also considerably larger in patients. However, due to the lifelong stability of fingerprints, effects of age would not have to be expected, except for a larger number of scars and wounds accumulated through external life events. It is assumed, though, that these had been minimized by the exhaustive quality control and improvements carried out when preprocessing the images.

There are also other issues to bear in mind to have a rational idea of the practical implications of our study. First of all, while we have achieved a reasonably high level of accuracy, applicability of the predictive models should be eventually measured through their positive predictive value (PPV, ie, the probability that an individual classified as patient by the model is really a patient) instead of the accuracy (ie, the probability of an individual to be classified in his/her correct class). Apart from accuracy, the PPV depends on the prevalence of the disorder. A low prevalence such as the one for schizophrenia, which is approximately 1% of the general population, leads to very low PPVs even for high accuracies. This restricts the practical utility of the developed algorithms to high risk subpopulations such as groups of individuals with prodromal symptoms or at high genetic risk for schizophrenia. In the former group (which may also include individuals with a possible first episode of psychosis) the predictive algorithms could become a valuable clinical tool if used adequately and with results cautiously interpreted. However, for individuals at high genetic risk for schizophrenia, clear ethical guidelines would have to be put in place to regulate their use, especially in the early stages of life.

From the design and results of our study, we cannot derive any conclusion on the specificity of the abnormalities

found. Indeed, prenatal stress has also been considered a risk-factor for other psychiatric disorders (see reviews by Lautaresku<sup>30</sup> and Van den Bergh<sup>31</sup>). However, previous studies comparing fingerprints in schizophrenia with those from other psychiatric disorders such as bipolar disorder, posttraumatic stress disorder or anxiety, have reported high levels of specificity in schizophrenia.<sup>32,33</sup> Still, these studies were based on small sample sizes and cannot be considered conclusive. Also, if the observed abnormalities are indicative of the intensity and extent of insults occurring during gestation, they may be taken as a proxy for the impact of prenatal stress, especially when information on prenatal stress had not been properly registered in clinical records and it is instead unreliably recalled. Although the primary objective of our study was to evaluate the predictive power of CNNs in schizophrenia, additional outcomes such as the heatmaps may have a broader interest. Interindividual variability in the extent and location of the anatomical patterns observed in these heatmaps is probably related to the well-known between subject variability in the presence and location of arches, loops, and whorls in fingerprints. In that sense, the convolutional networks used here are ideal for fingerprint characterization, as they rely on space invariant filters that are able to detect finger patterns regardless of their position in the image.

In spite of research in the field spanning several decades now, placing our results in relation to previous findings may be difficult as no clear consensus has been achieved yet. Maybe, the significant difference in total finger ridge count reported in the meta-analysis by Golembo-Smith et al.<sup>4</sup> could be linked to our findings, although such measurement is anatomically unspecific. While a deeper understanding of patterns in heatmaps and their relation to schizophrenia warrants future work, other areas of interest include the development of models for differential diagnosis between affective and non-affective psychoses, for disorder prognosis (eg, prediction of long-term clinical evolution) and the design of multi-input models combining fingerprint images with other sources of data.

To the best of our knowledge, this is the first study analyzing the potential utility of fingerprint images to automatically diagnose schizophrenia through DL, and our results support the feasibility of such an approach. Moreover, the lifelong stability of fingerprints also supports their potential value as predictors of risk of psychosis, especially if combined with other sources of data and if applied to high prevalence subpopulations like those of individuals presenting prodromal psychotic symptoms or subjects with significant genetic risk for schizophrenia.

### Supplementary Material

Supplementary material is available at <https://academic.oup.com/schizophreniabulletin/>.



## Acknowledgments

We thank Miguel Lechón for developing the software for data collection. This work was supported by several grants funded by the Instituto de Salud Carlos III and the Spanish Ministry of Science and Innovation (co-funded by the European Regional Development Fund/European Social Fund “Investing in your future”): Miguel Servet Research Contract (CP113/00018 to RS, CP116/00018 to EP-C, CP20/00072 to MF-V), PFIS Contract (FI19/0352 to MG-R). Research Mobility programme (MV18/00054 to EP-C), Research Projects (PI18/00877 and PI21/00525 to RS). It has also been supported by the Centro de Investigación Biomédica en Red de Salud Mental and the Generalitat de Catalunya: 2014SGR1573 and 2017SGR1365 to EP-C and SLT008/18/00206 to IF-R from the Departament de Salut. The authors have declared that there are no conflicts of interest in relation to the subject of this study.

## References

- Murray RM, Lewis SW. Is schizophrenia a neurodevelopmental disorder? *Br Med J (Clin Res Ed)*. 1988;296:63.
- Weinberger D. From neuropathology to neurodevelopment. *Lancet*. 1995;346:552–557.
- Fatjó-Vilas M, Gourion D, Campanera S, et al. New evidences of gene and environment interactions affecting prenatal neurodevelopment in schizophrenia-spectrum disorders: a family dermatoglyphic study. *Schizophr Res*. 2008;103:209–217.
- Golembo-Smith S, Walder DJ, Daly MP, et al. The presentation of dermatoglyphic abnormalities in schizophrenia: a meta-analytic review. *Schizophr Res*. 2012;142:1–11.
- Bracha HS, Torrey EF, Bigelow LB, Lohr JB, Linington BB. Subtle signs of prenatal maldevelopment of the hand ectoderm in schizophrenia: a preliminary monozygotic twin study. *Biol Psychiatry*. 1991;30:719–725.
- Holt SB. *The Genetics of Dermal Ridges*. Thomas Springfield IL; 1968.
- Karmakar B, Yakovenko K, Kobylansky E. Complex segregation analysis of quantitative dermatoglyphic traits in five Indian populations. *Ann Hum Biol*. 2005;32:445–468.
- Reed T, Opitz JM. Dermatoglyphics in medicine—problems and use in suspected chromosome abnormalities. *Am J Med Genet*. 1981;8:411–429.
- Schaumann B, Alter M. *Dermatoglyphics in Medical Disorders*. Springer-Verlag; 1976.
- Babler WJ. Embryologic development of epidermal ridges and their configurations. *Birth Defects Orig Artic Ser*. 1991;27:95–112.
- King S, Mancini-Marie A, Brunet A, Walker E, Meaney MJ, Laplante DP. Prenatal maternal stress from a natural disaster predicts dermatoglyphic asymmetry in humans. *Dev Psychopathol*. 2009;21:343–353.
- Cohen-Bendahan CC, van de Beek C, Berenbaum SA. Prenatal sex hormone effects on child and adult sex-typed behavior: methods and findings. *Neurosci Biobehav Rev*. 2005;29:353–384.
- Bramon E, Walshe M, McDonald C, et al. Dermatoglyphics and schizophrenia: a meta-analysis and investigation of the impact of obstetric complications upon a–b ridge count. *Schizophr Res*. 2005;75:399–404.
- Van Os J, Woodruff PWR, Fañanas L, et al. Association between cerebral structural abnormalities and dermatoglyphic ridge counts in schizophrenia. *Compr Psychiatry*. 2000;41:380–384.
- Reilly JL, Murphy PT, Byrne M, et al. Dermatoglyphic fluctuating asymmetry and atypical handedness in schizophrenia. *Schizophr Res*. 2001;50:159–168.
- Goodfellow I, Bengio Y, Courville A. *Deep Learning*. MIT Press; 2016.
- Schmidhuber J. Deep learning in neural networks: an overview. *Neural Networks*. 2015;61:85–117.
- Shen D, Wu G, Suk HI. Deep learning in medical image analysis. *Annu Rev Biomed Eng*. 2017;19:221–248.
- Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60–88.
- Decuyper M, Maebe J, Van Holen R. Artificial intelligence with deep learning in nuclear medicine and radiology. *EJNMMI Phys*. 2021;8:8–81.
- Hong L, Wan Y, Jain A. Fingerprint image enhancement: algorithm and performance evaluation. *IEEE Trans Pattern Anal Mac Intell*. 1998;20:777–89.
- Chollet F, Allaire JJ. *Deep Learning with R*. Manning Publications; 2018.
- Del Ser T, González-Montalvo JI, Martínez-Espinosa S, Delgado-Villalpalos C, Bermejo F. Estimation of premorbid intelligence in Spanish people with the Word Accentuation Test and its application to the diagnosis of dementia. *Brain Cogn*. 1997;33(3):343–356.
- Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Proc IEEE Int Conf Comput Vis (ICCV)*. 2017:618–626.
- Ripke S, Neale BM, Corvin A, et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511:421–427.
- Legge SE, Santoro ML, Periyasamy S, Okewole A, Arsalan A, Kowalec K. Genetic architecture of schizophrenia: a review of major advancements. *Psychol Med*. 2021;51(13):2168–2177.
- van Erp TG, Walton E, Hibar D, et al. Cortical abnormalities in 4474 individuals with schizophrenia and 5098 controls via the ENIGMA consortium. *Biol Psychiatry*. 2018;84(9):644–654.
- Gutman BA, van Erp TG, Alpert K, et al. A meta-analysis of deep brain structural shape and asymmetry abnormalities in 2,833 individuals with schizophrenia compared with 3,929 healthy volunteers via the ENIGMA Consortium. *Hum Brain Mapp*. 2022;43(1):352–372.
- Jauhar S, Johnstone M, McKenna PJ. Schizophrenia. *Lancet*. 2022;399:473–486.
- Lautarescu A, Craig MC, Glover V. Prenatal stress: effects on fetal and child brain development. *Int Rev Neurobiol*. 2020;150:17–40.
- Van den Bergh BRH, van den Heuvel MI, Lahti M, et al. Prenatal developmental origins of behavior and mental health: the influence of maternal stress in pregnancy. *Neurosci Biobehav Rev*. 2020;117:26–64.
- Domany Y, Levy A, Cassan SM, et al. Clinical utility of biomarkers of the hand in the diagnosis of schizophrenia. *Psychiatry Res*. 2018;260:105–110.
- Zvi Shamir E, Levy A, Morris Cassan S, Lifshitz T, Shefler G, Tarrasch R. Do biometric parameters of the hand differentiate schizophrenia from other psychiatric disorders? A comparative evaluation using three mental health modules. *Psychiatry Res*. 2015;228(3):425–430.