UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Medicine and Life Sciences
School for Life Sciences
Master of Biomedical Sciences

Master's thesis

Performance of a standalone photoplethysmography-based algorithm that detects atrial fibrillation: analysis of real-world data

Jente Gieghase

Thesis presented in fulfillment of the requirements for the degree of Master of Biomedical Sciences, specialization Bioelectronics and Nanotechnology

SUPERVISOR :
Prof. dr. Lars GRIETEN

MENTOR :
MSc Stijn EVENS
Ir. Glenn DE WITTE

2020
2021

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Medicine and Life Sciences
School for Life Sciences
Master of Biomedical Sciences

Master's thesis

Performance of a standalone photoplethysmography-based algorithm that detects atrial fibrillation: analysis of real-world data

Jente Gieghase

Thesis presented in fulfillment of the requirements for the degree of Master of Biomedical Sciences, specialization Bioelectronics and Nanotechnology

SUPERVISOR :
Prof. dr. Lars GRIETEN

MENTOR :
MSc Stijn EVENS
Ir. Glenn DE WITTE

Performance of a standalone photoplethysmography-based algorithm that detects atrial fibrillation: analysis of real-world data*

J. Gieghase[1], S. Evens[2], G. De Witte[2] , L. Grieten[1,2]

[1]Department of Medicine and Life Sciences, University of Hasselt, Martelarenlaan 42 - B - 3500 Hasselt
[2]Qompium NV, Kempische Steenweg 303/27 - B - 3500 Hasselt

*Running title: *Performance of a PPG-based AF detection algorithm*

To whom correspondence should be addressed:
Lars Grieten, Tel: +32 89 32 15 12; Email: lars.grieten@uhasselt.be

Keywords: atrial fibrillation, photoplethysmography, algorithm, diagnostic performance, real-world data

ABSTRACT

Background – Photoplethysmography (PPG) deriving applications may facilitate timely detection of atrial fibrillation (AF). However, some questions concerning the real-world performance of proprietary AF detection algorithms remain to be answered.

Objective – To assess the diagnostic performance of FibriCheck's proprietary AF detection algorithm by retrospective analysis of real-world data.

Methods – Anonymous user data was extracted from the secured FibriCheck system. Every algorithm-based result was validated by trained medical technicians under the supervision of a cardiologist. Measurements that were either unrevised or revised as insufficient quality were excluded. The remaining measurements were classified as either non-AF or possible AF. Dichotomous comparison between the results enabled the determination of the performance on measurement and participant level. Possible differences in performance between different categories according to age, CHA_2DS_2-VASc score, and AF history were investigated.

Results – In total, 409,299 validated measurements derived from 2,163 participants were included in the analysis. The algorithm yielded 98.7% sensitivity and 98.5% specificity on measurement level, while the analysis on participant level resulted in 99.0% sensitivity and 80.3% specificity. Similar results were obtained for each analysis according to age, CHA_2DS_2-VASc score, and AF history but as the AF prevalence and the median number of measurements per participant increased, so did the number of participants who received correct possible AF results on participant level.

Conclusion – Although the algorithm was demonstrated to be an effective rule-out solution, human revision remains essential for specific AF detection. Further improvements of the algorithm might contribute to its implementation as a standalone solution in the real-world setting.

INTRODUCTION

The heart's electrical activity

The cardiac conduction system (CCS) controls the generation and propagation of electrical impulses and hereby coordinates rhythmic contractions of the atria and ventricles (Figure 1) (1-3). The electrical impulses are usually generated by pacemaker cells located within the sinoatrial node and are subsequently propagated throughout the atria towards the atrioventricular node (AVN) (2, 3). Normally, the AVN is the only conductive tissue that connects the CCS between the atria and the ventricles. Furthermore, the AVN induces a delay in impulse propagation, which allows the ventricles to fill properly before their mechanical contraction (2). Next, the electrical impulse is rapidly conducted throughout the ventricles via the bundle of His, which is divided into left and right bundle branches. At last, the Purkinje fiber network is electrically activated, causing ventricular contraction which ejects the blood into the aorta (1, 3). The well-coordinated generation and propagation of electrical impulses by the CCS are essential to maintain rhythmic contractions that ensure proper cardiac function. Any impairment of the CCS may result in heart rhythm disorders, including atrial fibrillation (AF) (2).

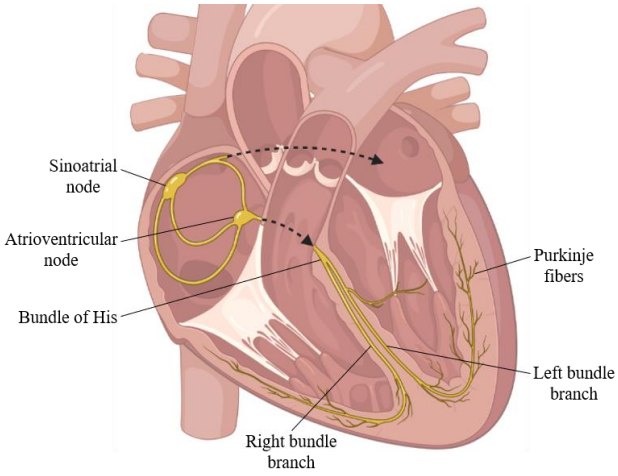

Figure 1: The cardiac conduction system. Electrical impulses are initiated in the sinoatrial node and propagate along the atrioventricular node towards the bundle of His and its branches. At last, the terminal Purkinje fibers are electrically activated and contraction occurs.

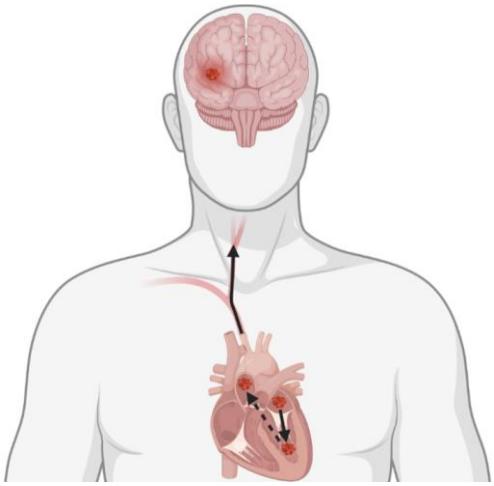

Figure 2: Development of an AF-related stroke. Atrial fibrillation may lead to blood stasis and hereby promote thrombus formation. A dislodged thrombus may travel from the atrium towards the brain where it can occlude arterial blood vessels, causing a cardioembolic stroke.

Importance of screening for atrial fibrillation

AF is the most common sustained cardiac arrhythmia with an estimated worldwide prevalence between 2 and 4% (4). Moreover, the AF prevalence is expected to rise a 2.3-fold in the coming decades, due to the aging population and the intensifying search for undiagnosed patients (5, 6). Apart from the age, there are additional risk factors that may contribute to the development of AF, including hypertension, heart failure, and coronary artery disease. (7-10) AF is characterized by uncoordinated electrical activation of the atria, causing ineffective atrial contractions (i.e., fibrillation) (10). This may lead to impaired blood flow or blood stasis which can result in the formation of a thrombus. The dislodgement of such a thrombus can cause a cardioembolic stroke (Figure 2) (11, 12). Therefore, AF causes an approximate 3- to 5-fold increase in stroke risk (13). AF-related strokes are usually severe, highly recurrent, and often result in death or permanent disability (14-16). Apart from strokes, additional AF-related outcomes such as heart failure, increased hospitalization rate, and a 1.5- to 3.5-fold increased mortality rate may also impair the patient's quality of life (10). Not only does AF have a profound impact on the patient's life, but it also places a significant burden on societal health and the health economy (10).

AF is generally considered as a progressive disease (17). The progression of AF is associated with adverse cardiovascular events and is attributed to atrial remodeling caused by various pathophysiological mechanisms (10, 18, 19). Although some AF patients may experience symptoms such as palpitations and dyspnea, more than 50% of the patients are initially asymptomatic (10, 20). Many of these asymptomatic AF patients will only seek help when symptoms or complications occur. This may lead to a delayed diagnosis by which AF progression has advanced already (10). Due to AF's often infrequent and asymptomatic nature, more than 25% of AF patients receive their diagnosis only after an AF-related stroke has occurred (21). However, timely AF detection allows for early intervention that may interrupt progression and reduce the stroke risk, thereby improving patient outcomes (17, 21). Therefore, the European Society of Cardiology (ESC) recommends opportunistic screening for AF by pulse taking or electrocardiogram (ECG) strip in patients from the age of 65 years (10). Not only will opportunistic screening facilitate timely AF detection in patients from the age of 65, but it is also more cost-effective than routine practice and systematic screening (22, 23).

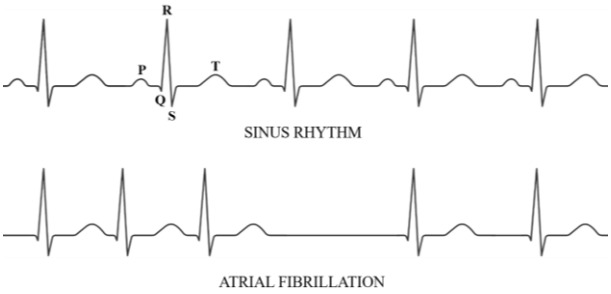

Figure 3: ECG comparison between sinus rhythm and AF. The sinus rhythm shows discernable P, QRS complexes and T waves, while the AF heart rhythm displays no discernable P waves and irregular RR intervals.

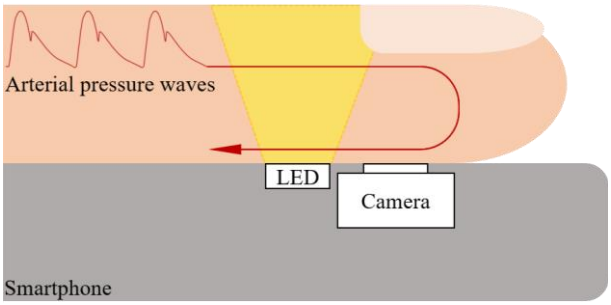

Figure 4: Smartphone-based PPG signal acquisition. Arterial blood pressure waves result in volumetric changes within the dense capillary beds of the fingertip. This process can be studied by using a smartphone's flashlight to illuminate the capillaries while the build-in camera measures the amount of reflected light.

Detection of AF via electrocardiography

Conventionally, the cardiologist's interpretation of an ECG is required to diagnose AF and a traditional 12-lead ECG is hereby considered as gold standard (24-26). Usually, an ECG consists of discernable P waves, QRS complexes, and T waves representing atrial depolarization, ventricular depolarization, and ventricular repolarization, respectively (Figure 3) (24). The ESC states that clinical AF should be diagnosed via an ECG without discernible repeating P waves and with irregular RR intervals (RRIs) (10, 26). The sometimes brief and infrequent nature of AF complicates the diagnosis via the conventional 10-second ECG strip, hereby emphasizing the demand for more convenient AF screening tools (26, 27). To facilitate the detection of paroxysmal AF, innovative ECG technologies including patches, wearables, and handheld devices were developed (28). Some of these technologies are based on single-lead ECG and have been proven to be reliable with respect to 12-lead ECG as gold standard (28). The usage of these ECG technologies may facilitate remote heart rhythm monitoring and AF detection at relatively low costs and efforts compared to the traditional 12-lead ECG approach (29, 30). Despite the advantages of portable ECG technologies, their availability and scalability are subject to a major limitation caused by the need for additional, specific, and/or costly hardware (31). Evidently, there is a demand for innovative technologies that can overcome these hurdles.

Detection of AF via photoplethysmography

Ejection of oxygen-rich blood into the aorta causes arterial blood pressure waves and results in blood volume pulse variation (Figure 4) (31, 32). These volumetric changes are the result of blood pooling in the capillaries during systole and their subsequent relaxation during diastole (33). It is possible to study these volumetric changes by using an optical technique called photoplethysmography (PPG) (32-34). Traditional PPG systems consist of a light-emitting diode (LED) as a light source and a photodetector that can detect the amount of light that is reflected by dense capillary beds through the skin (32, 33). Hence, the built-in LED flashlight and camera or photodetector of a smartphone or smartwatch device enable the measurement of PPG signals without any additional hardware (33).

Qompium NV (Hasselt, Belgium) developed the FibriCheck app that uses PPG technology to monitor heart rate and rhythm. This app is widely available on smartphones and several smartwatch devices and received regulatory approval by the Conformité Européenne (CE), the Food and Drug Administration (FDA), and the Therapeutic Goods Administration (TGA) (31, 34). Vandenberk et al. demonstrated the ability to use the app for heart rhythm detection by comparing peak-to-peak intervals (PPIs) and RRIs of simultaneously recorded PPG and single-lead ECG data (Figure 5) (33). The reported correlation of 99% between the corresponding PPIs and RRIs indicates that PPG could be used for heart rhythm monitoring (31, 33). Proesmans et al. validated the proprietary PPG-based AF detection algorithm for smartphone use in

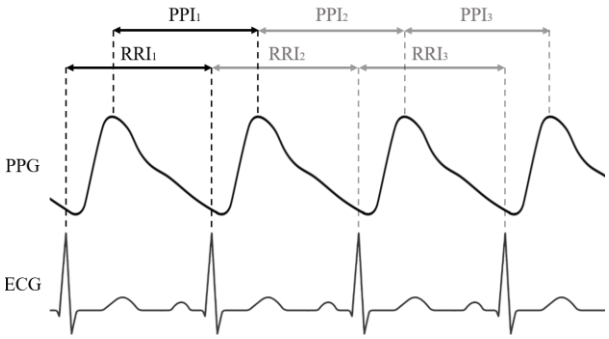

Figure 5: Comparison between ECG and PPG. Heart rhythm monitoring based on PPG is possible because the PPIs correspond to the RRIs from a simultaneously recorded ECG. The PPG signal displays the amount of reflected light and is conventionally inverted.

a primary care population. They reported 96% sensitivity and 97% specificity on participant level compared to traditional 12-lead ECG diagnosis by a cardiologist (31). In addition, Selder et al. indicated that the proprietary algorithm was able to detect AF based on wristband-derived PPG data and hereby obtained 100% sensitivity and 96% specificity on participant level by using a cardiologist's diagnosis based on single-lead ECG as reference (35). Moreover, Verbrugge et al. reported that the use of the FibriCheck smartphone app is feasible and attractive for mass screening to detect AF (34). There is no need for any additional hardware and as the use of smartphones and smartwatch devices continues to increase, this app shows potential as a convenient, cost-effective, and non-invasive method for heart rhythm monitoring (31, 33, 34, 36).

Research question

The performance of FibriCheck's AF detection algorithm has been thoroughly validated in various clinical settings (31, 37, 38). These studies recruited patients from primary and secondary care facilities, resulting in controlled patient populations. Apart from FibriCheck's use in a controlled clinical setting, the app is also used in the real-world setting. This free-living setting results in a heterogeneous population with adults of all ages with various health conditions and comorbidities. Moreover, there is limited control over the user population and the time period in which data can be collected. This research was conceived because the real-world performance of the proprietary AF detection algorithm and its impact on the users has not yet been thoroughly investigated. In this study, we determine the performance of FibriCheck's proprietary AF detection algorithm by analysis of real-world data with respect to offline human revision as gold standard. This study is not only an important addition to FibriCheck's transparency, but may also provide novel insights into the real-world performance and hereby lead to further improvement of the app and ultimately, better AF detection.

EXPERIMENTAL PROCEDURES

FibriCheck as data source

Every FibriCheck user has to create an account and is hereby required to register age, gender, and country of residence. In addition, the users can voluntarily complete an in-app questionnaire at any time after their first measurement. This survey is used to collect additional information including medication use and possible comorbidities that enable an estimation of the stroke risk via the CHA_2DS_2-VASc (i.e., congestive heart failure, hypertension, age $\geq$ 75 years, diabetes mellitus, previous stroke, vascular disease, age 65-75 years and female sex) score. PPG signals are obtained during one-minute measurements by placing a fingertip on the smartphone or by applying a smartwatch device to the wrist (34). Each measurement is automatically analyzed by the proprietary AF detection algorithm. Users with paid premium subscriptions are able to perform an unlimited number of measurements that receive a secondary offline review by trained medical technicians under the supervision of a cardiologist within 48 hours. The duration of a subscription can vary from months to years and every user is advised to measure at least twice a day or when they experience symptoms. The exact number of measurements performed by a user is dependent on both duration of their subscription and their measurement behavior (i.e., compliance and motivation). Instructions to obtain a high-quality PPG signal are displayed on-screen before the first measurement takes place. It is preferred that users sit down with both arms resting on a firm surface, holding the smartphone vertically in their hand and placing the index finger of the other hand near the flashlight and on top of the backside camera, without putting firm pressure. Following these instructions is important to obtain high-quality PPG signals. However, additional factors including cold

hands and callus formation can also decrease the obtained signal quality. After each measurement, information about the user's activity and symptoms is requested.

Data collection

Anonymous premium user data from a period of 16 months (January 2020 - April 2021) was extracted from the secured FibriCheck cloud. During this time window, 4 versions (1.3.1 - 1.3.4) of the proprietary algorithm were subsequently in production. This study complies with the Declaration of Helsinki and each participant granted permission to process their anonymized personal data for scientific and research purposes by accepting the Privacy Policy and the company's Terms of Service before account activation.

PPG data processing

Each PPG measurement is automatically analyzed by the highly sensitive proprietary AF detection algorithm. This machine-learning algorithm consists mainly of a deep neural network and assigns each measurement to one of the following categories: insufficient quality (blue, quality), regular heart rhythm (green, normal), possible non-AF irregularity (orange, warning), and possible AF (red, urgent). The non-AF irregularities include bradycardia (episodes), tachycardia (episodes), and frequent extrasystoles. After each measurement, the user receives immediate feedback from the algorithm. These results include the heart rate, heart rhythm, and a detailed measurement report with the PPG output and corresponding advice which can easily be shared with their physician. If the obtained measurements are of insufficient quality, instructions on how to enhance the quality during the next measurement are displayed. In case of a regular heart rhythm, the app indicates that there is no action required. For possible non-AF irregularities, the app informs the users that there is no immediate action required but does advise them to show their results to their treating physician, during the next scheduled visit. When a possible AF measurements is detected, the app indicates that immediate action is required and advises the user to consult their physician for further examination. The initial results from the algorithm's automated analysis of premium data are either confirmed or corrected by the human revision that is included in FibriCheck's premium services. For every raw PPG signal, a tachogram and Poincaré plot are created to facilitate human revision (Figure 6). The tachogram shows the time between successive PPIs in milliseconds, while the Poincaré visualizes the randomness of the heart rhythm by plotting each PPI relative to the previous PPI.

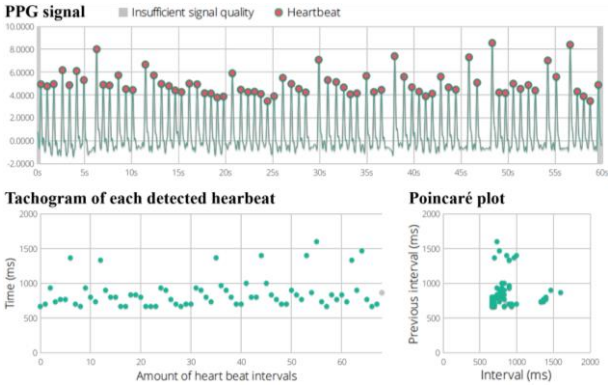

Figure 6: Measurement-derived data. Every measurement results in a raw PPG signal, tachogram and Poincaré plot. This particular measurement displays frequent ectopic beats and was classified as non-AF.

Analysis of the diagnostic performance

All measurements indicated as insufficient quality by the human revision were excluded from the analysis, along with all measurements that weren't revised yet. Measurements reviewed as insufficient quality by the algorithm, regular heart rhythms, and possible non-AF irregularities were classified as non-AF measurements for a dichotomous comparison with the possible AF measurements. Revision of the PPG signal by trained medical technicians under the supervision of a cardiologist was used as gold standard to determine the algorithm's diagnostic performance to detect AF. Diagnostic metrics including sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV) were calculated to determine the proprietary algorithm's performance to detect AF. Calculations for the PPV and NPV were based on the AF prevalence within the corresponding population.

The algorithm's diagnostic performance was determined on measurement level by analysis of all human revised measurements. In addition, the performance on participant level was determined to gain more insight into the possible impact of the algorithm's findings on the users. For this purpose, participants were classified according to the

following 4 categories. True positive users received at least one possible AF measurement, confirmed by the human revision. These users were correctly advised to consult a medical professional. The false-positive users received at least one possible AF measurement, and all the received possible AF measurements were corrected by the human revision as non-AF. In this case, the absence of the human revision would have led to the wrong advice to consult a medical professional. Users who received exclusively non-AF measurements confirmed by the human revision were indicated as true negative users. These users received the correct advice that no (immediate) action was required. At last, false-negative users received exclusively non-AF measurements, of which at least one was corrected by the human revision as possible AF. Without the human revision, these users would have been wrongly informed that there was no (immediate) action required.

Additional analyses were performed to determine possible differences in the algorithm's performance between different categories according to age, AF history, and CHA_2DS_2-VASc score that is used to estimate the stroke risk of AF patients. Participants were hereby classified according to 3 age groups (<40, 40-65, and >65), 2 AF history groups (no AF history and AF history), and 3 stroke risk groups (low, moderate, and high risk). For males, a CHA2DS2-VASc score of 0 indicated low risk, 1 indicated moderate risk, and 2 or greater was considered as high risk. Females followed a similar classification in which the threshold of every risk group was increased by one due to their sex as a risk factor. The groups according to AF history and CHA_2DS_2-VASc score exclusively consisted of participants who voluntarily provided additional data about these characteristics via an in-app medical questionnaire. Shapiro-Wilk tests were used to test if variables were normally distributed. Normally distributed continuous variables were expressed by means and standard deviations, otherwise, these variables were expressed by medians and interquartile ranges (IQR). Fisher's exact tests and Fisher-Freeman-Halton exact tests were used to determine if potential significant differences were present in the diagnostic metrics between groups. These statistical tests were 2-sided with a 5% significance level. RStudio (Version 1.3.1073) was used to perform statistical analyses.

RESULTS

Study population

A total of 2,168 users performed at least one measurement during the study period, resulting in a total of 429,270 measurements (Figure 7). In total, 19,971 (4.7%) measurements were excluded because the human revision was either pending (n = 228) or indicated insufficient quality (n = 19,743). This resulted in the exclusion of 5 participants for which no human revised measurements were present. As a result, the study population comprised 2,163 participants that performed 409,299 measurements (Table 1). Of these participants, 1,188 (54.9%) were men and the median age was 54 (IQR = 23) years. Additional information to determine the stroke risk was provided voluntarily by 2,085 (96.4%) participants. This portion of the study population included 508 (24.4%) participants who had been previously diagnosed with AF and resulted in a median CHA_2DS_2-VASc score for stroke risk of 1 (IQR = 2).

Table 1: Characteristics of the study population. Information about age and gender were mandatory and therefore provided by all 2,163 participants. Additional information was provided voluntarily by 2,085 (96.4%) participants via an in-app medical questionnaire.

	Total population (n = 2,163)
Mandatory data (n = 2,163)	
Age, years, median (IQR)	54 (23)
Male, n (%)	1,188 (54.9)
Voluntary data (n = 2,085)	
Hypertension, n (%)	602 (28.9)
Diabetes mellitus, n (%)	151 (7.2)
Vascular disease, n (%)	159 (7.6)
Heart failure, n (%)	150 (7.2)
Stroke (CVA / TIA), n (%)	113 (5.4)
Pacemaker, n (%)	40 (1.9)
Arrhythmias, n (%)	949 (45.5)
Atrial fibrillation, n (%)	508 (24.4)
Blood thinner, n (%)	611 (29.3)
CHA_2DS_2-VASc score, median (IQR)	1 (2)

Abbreviations: IQR, interquartile range. AF, atrial fibrillation. CVA, cerebrovascular accident. TIA, transient ischemic attack.

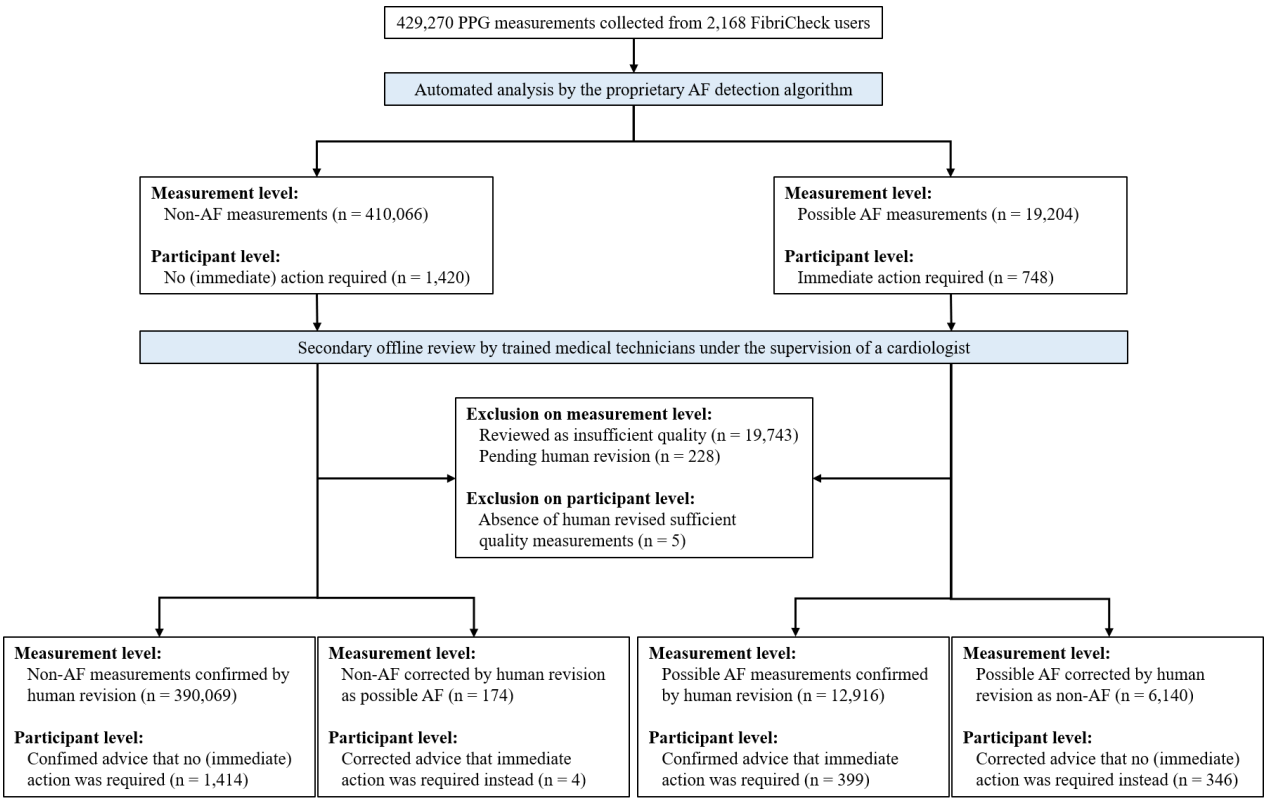

Figure 7: Study flowchart. The algorithm's results were classified as either non-AF or possible AF and communicated to the participants together with corresponding advice. Next, the human revision took place and measurements of insufficient quality and pending revisions were excluded along with the participants who did not perform a single human revised sufficient quality measurement. At last, the results and their corresponding advice were either corrected or confirmed by the human revision. PPG, photoplethysmography. AF, atrial fibrillation.

Analysis of real-world PPG data

Measurement level – The median number of measurements per participant was 59 (IQR = 143.5). In 402,985 (98.5%) measurements, the algorithm's result was confirmed by human revision (Supplementary Table 1). These correctly classified measurements included 12,916 true positives and 390,069 true negatives. The remaining 6,314 (1.5%) measurements were adjusted by human revision and consisted of 6,140 false positives and 174 false negatives. This resulted in a sensitivity and specificity of 98.7% and 98.5%, respectively (Figure 8). Moreover, the human revision revealed an AF prevalence of 3.2%. Based on the obtained sensitivity, specificity, and prevalence, the algorithm yielded a PPV and NPV of 67.8% and 100%, respectively.

Participant level – The algorithm's results matched the definite result after the human revision in 1,813 (83.8%) participants (Supplementary Table 1). Most participants received correct results as 1,414 (65.4%) and 399 (18.5%) participants were classified as true negatives and true positives, respectively. Apart from these correct results, 346 (16%) participants were classified as false positives and 4 (0.2%) as false negatives. These results yielded a sensitivity of 99.0% and specificity of 80.3%. Although the sensitivity was similar compared to the measurement level, the specificity was relatively low (Figure 8). The medical technicians identified 18.6% of the participants as possible AF patients. In addition to the specificity, the PPV also showed a substantial decrease from 67.8% on measurement to 53.6% on participant level. The NPV was 99.7% and hereby remained similar to the measurement level.

Diagnostic metrics per age group

Measurement level – The participants younger than 40 years performed 70,201 (17.2%) measurements. The 40 to 65 year-olds performed 220,159 (53.8%) measurements and 118,939 (29.1%) measurements were performed by

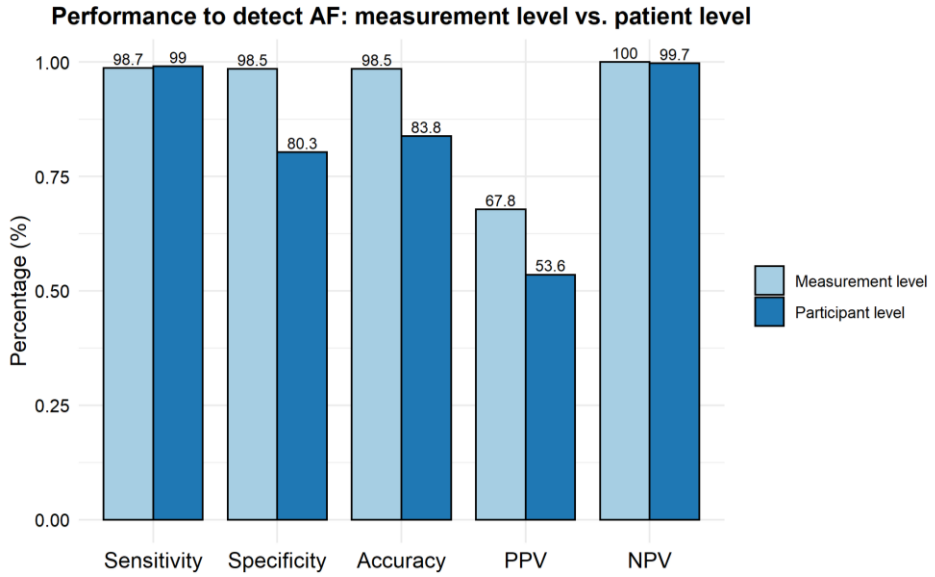

Figure 8: Diagnostic performance on measurement level (n = 409,299) and participant level (n = 2,163) that reveal the algorithm's performance to detect atrial fibrillation (AF) in a real-word population. Apart from the corresponding sensitivity and specificity, the AF prevalence on measurement level (3.2%) and on participant level (18.6%) were used to calculate the positive predictive value (PPV) and negative predictive value (NPV).

participants older than 65 years (Supplementary Table 2). The sensitivities obtained for these age groups were 97.6%, 98.4%, and 99.1%, respectively, and their specificities 98.1%, 98.9%, and 97.9%, respectively. These results suggest that the sensitivity increases along with the age and that the highest specificity is obtained for 40-65 year-olds. The sensitivity for participants older than 65 was significantly higher than for 40 to 65 year-olds ($P < 0.001$) and participants younger than 40 ($P = 0.028$). In addition, the sensitivity of <40 and 40-65 age groups did not differ significantly ($P = 0.349$). The specificity for 40 to 65 year-olds was significantly higher than for participants younger than 40 ($P < 0.001$) and participants above 65 years old ($P < 0.001$). At last, the specificity for the <40 age group was significantly lower ($P < 0.001$) than for the >60 age group.

Participant level – This analysis included 495 (22.9%) participants younger than 40 years along with 1,166 (53.9%) 40 to 65 year-olds and 502 (23.2%) participants older than 65 years (Supplementary Table 3). The median number of measurements per participant (85.5, IQR = 236.3) was higher for participants older than 65 than for those from 40 to 65 year-olds (56, IQR = 129) and participants younger than 40 (39, IQR = 104.5). The following sensitivities were obtained arranged from the youngest to the oldest age group: 97.6%, 98.4% and 100% (Figure 9). This trend is similar to the one found on the measurement level. The highest specificity of 83.3% was obtained for the 40 to 65 year-olds, while the lowest specificity of 74.8% was obtained for the youngest age group (Supplementary Table 4). The obtained PPV values increase towards the older age groups from 26.4% to 53,3% and 71.4%. The lowest NPV value of 99.6% was calculated for participants between 40 and 65 years old.

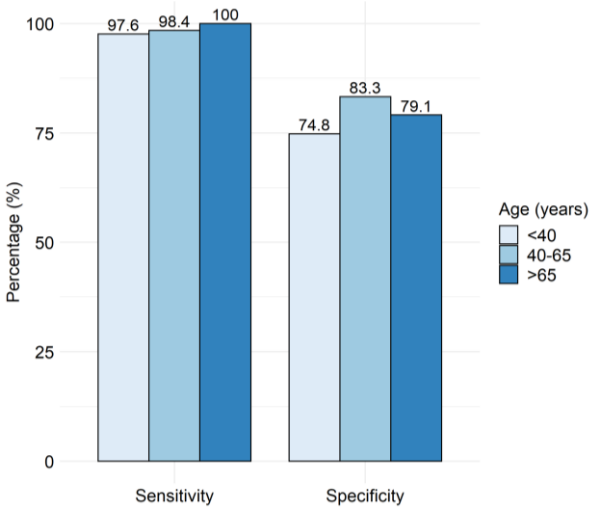

Figure 9: Sensitivity and specificity on participant level per age group. 2,136 participants were classified in 3 groups according to age in years: <40 (n = 495), 40-65 (n = 1,166) and >65 (n = 502).

Diagnostic metrics per risk group

Measurement level – The largest portion of measurements was performed by participants at low risk and accounted for 160,038 (40.8%) measurements (Supplementary Table 5). The moderate-risk participants performed the smallest portion with a total of 104,299 (26.6%) measurements. Those at high risk performed the remaining 128,115 (32.6%) measurements. The sensitivity increased gradually along with the risk group from 98.3% at low risk to 98.9% at moderate risk, and 99.1% at high risk. The specificity reached its highest value of 98.9% in the moderate-risk group, followed by the low-risk group with 98.3% and the high-risk group with 98.1%. Although there was no significant difference ($P = 0.439$) between the sensitivity for the high and moderate risk groups, the sensitivity for low-risk participants was significantly lower than for the moderate risk ($P = 0.047$) and high-risk groups ($P < 0.001$). The specificity obtained for the moderate risk group was significantly higher than for the low risk ($P = 0.002$) and moderate risk group ($P < 0.001$). At last, the specificity for high-risk participants was also significantly lower ($P < 0.001$) than for low-risk participants.

Participant level – A total of 1,105 (53%) participants were at low risk and the remaining 980 (47%) participants were almost equally spread among moderate and high-risk participants (Supplementary Table 6). The median number of measurements per participant increased from 53 (IQR = 121) for low risk to 62 (IQR = 160) for moderate-risk and reached 83 (IQR = 214) for the high-risk group. The sensitivity reached its highest value of 99.4% for the high-risk participants and as the sensitivity for the moderate risk group was at its lowest with 99.0%, there was no gradual increase along with the risk present (Figure 10). However, the specificity did increase along with the risk from 79.0% for low-risk to 80.3% for moderate-risk and 83.0% for high-risk groups. As the CHA_2DS_2-VASc score increased, so did the AF prevalence from 12.2% to 20.7% and 32.1%, respectively (Supplementary Table 7). The obtained PPV values increased substantially towards the high-risk group from 39.7% to 73.4%, while the NPV values of these groups decreased slightly from 99.9% to 99.6%.

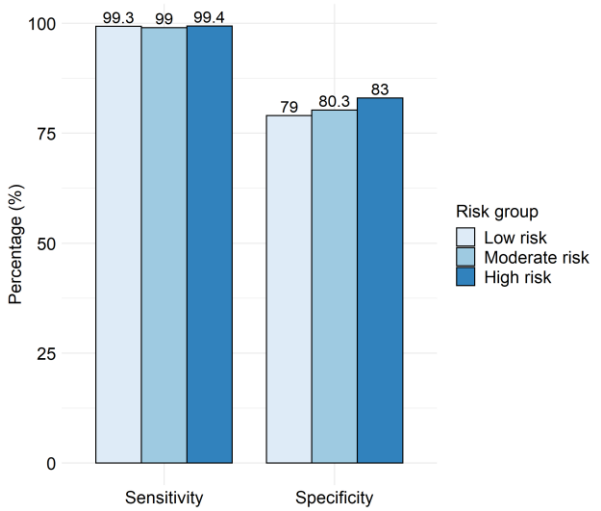

Figure 10: Sensitivity and specificity on participant level per risk group. 2,085 participants were classified in 3 groups according to their CHA_2DS_2-VASc score: Low (n = 1,105), moderate (n = 487) and high risk (n = 493).

Diagnostic metrics per AF history group

Measurement level – A total of 148,792 (37.9%) measurements were performed by the participants with a known history of AF. (Supplementary Table 8). The sensitivity for participants with a history of AF (99.0%) was higher than for those without previously diagnosed AF (98.6%). On the contrary, the specificity was lower for participants with a history of AF (98.2%) than for those without (98.5%). Fisher's exact tests revealed that both sensitivity ($P = 0.025$) and specificity ($P < 0.001$) differed significantly between these 2 groups.

Participant level – In total, 508 (24.4%) participants indicated that they were previously diagnosed with AF (Supplementary Table 9). These participants with a known history of AF obtained 105 (IQR = 211.3) as the median number of measurements per participant. This was substantially higher than the median of 51 (IQR = 125) obtained for participants without AF history. Both sensitivity (100%) and specificity (80.2%) were the highest for the participants without AF history (Figure 11). The sensitivity and specificity for participants with AF history yielded 98.7% and 79.5%, respectively. The group with AF history obtained the highest prevalence of 44.3% (Supplementary Table 10).

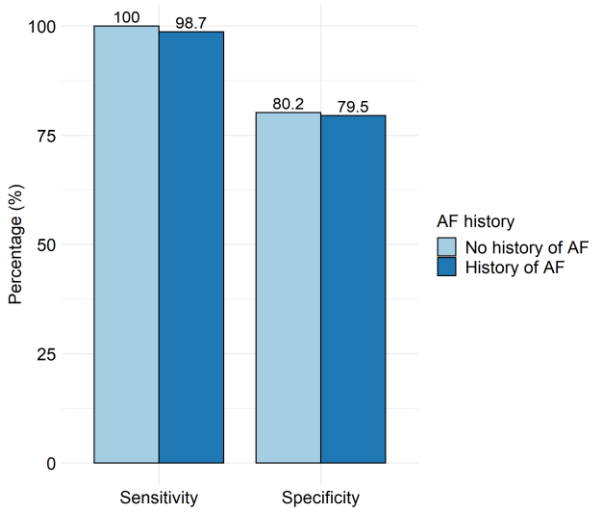

Figure 11: Sensitivity and specificity on participant level per atrial fibrillation (AF) history group. 2,085 participants were classified in 2 groups according to their history of AF: No history of AF (n = 1,577) and history of AF (n = 508)

DISCUSSION

The real-world diagnostic performance

In this study, the performance of FibriCheck's AF detection algorithm was determined by retrospective analysis of real-world data. It is therefore important to note that the results obtained during this study refer to the performance of the algorithm as a standalone solution (i.e., if we were to exclude the human revision). Therefore, the reported results may not be confused with the performance of the entire FibriCheck application, which includes both an analysis by the algorithm and a revision by medical technicians under the supervision of a cardiologist.

The overall sensitivity on measurement level was 98.7% and increased slightly to 99.0% on participant level (Figure 8). In addition, the NPV was 100% on measurement and decreased slightly to 99.7% on participant level. These results confirm that the algorithm is highly sensitive for the detection of AF. As a result, only few measurements caused a false negative result, which indicates that the algorithm is a reliable tool to rule out the presence of AF in a real-world population. However, the algorithm's specificity decreased from 98.5% on measurement level to 80.3% on participant level and a similar trend was seen for the PPV of 67.8% on measurement level and 53.6% on participant level. On both levels, the sensitivity was higher than the specificity and the NPV was higher than the PPV. These results indicate that the algorithm is less specific than it is sensitive to detect AF. Although it was indicated that the algorithm is highly sensitive for AF detection, further improvement regarding the specificity is required to enable its use without additional human revision. This was confirmed by comparison of the 53.6% PPV that was obtained on participant level with the 71% PPV that was reported by the Apple Heart study in which a real-world population was monitored using the Apple Watch PPG sensor and received an ECG patch analysis to diagnose possible AF upon the detection of irregular tachograms (39).

Although FibriCheck's proprietary algorithm may not offer a standalone solution to detect AF yet, the combination of the algorithm and a human revision does enable the use of the FibriCheck application as a reliable and convenient tool to detect AF. Therefore, these results highlight the importance of human revision by FibriCheck's trained medical technicians which has been proven reliable with 90.9 to 96.2% sensitivity and 97.8 to 98.8% specificity compared to 12-lead ECG diagnosis by a cardiologist. This human revision includes the raw PPG signal and is facilitated by the implementation of the tachogram and Poincaré plot. Apart from the trained medical technicians, Gruwez et al. indicated that physicians were also able to perform these PPG revisions with 95.5% sensitivity and 92.5% specificity compared to a 12-lead ECG diagnosis by a cardiologist and hereby achieved a similar diagnostic accuracy as for single-lead ECG measurements. (40). This highlights the potential of training physicians to revise the PPG signals for their patients and might enhance the accessibility of FibriCheck as a reliable tool for AF detection. Moreover, van der Velden et al. recently reported a structured stepwise practical guide on PPG signal interpretation which was developed based on presented experiences from TeleCheck-AF and will facilitate the implementation of PPG technology in the management of AF patients (41).

Additional analyses

The algorithm's performance per age, risk, and AF history group obtained results similar to the overall sensitivity, specificity, accuracy, and NPV on both the measurement and participant level. However, a

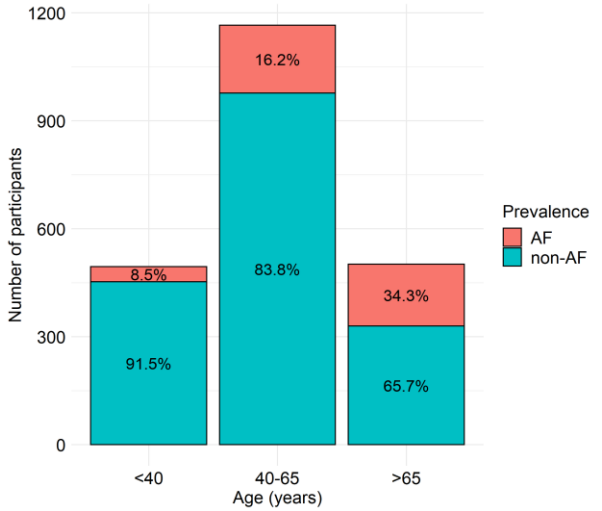

Figure 12: Prevalence of atrial fibrillation (AF) on participant level for each age group. 2,136 participants were classified in 3 groups according to their age.

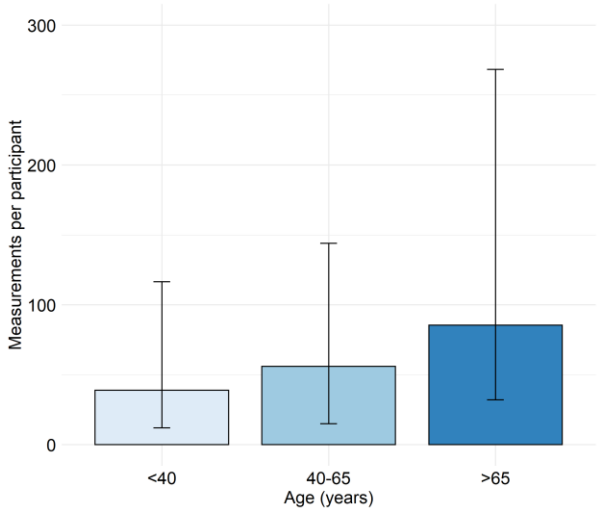

Figure 13: Median of measurements per participant for each age group. Interquartile ranges are indicated by the error bars.

clear trend was indicated by which the PPV on participant level increased together with the age, risk, and presence of AF history. For every additional analysis, this increase in PPV was associated with an increase in AF prevalence and the median number of measurements per person.

This trend was the most pronounced in the analysis of different age groups where the PPV values increased towards the older age groups from 26.4% for the <40 age group to 71.4% for the >65 age group. This 71.4% PPV value for participants older than 65 years is substantially higher than the 60% PPV that is reported by the Apple Heart study for participants from 65 years of age (39). This may indicate that the use of FibriCheck's algorithm as a standalone solution for AF detection is more reliable than the use of the Apple Watch PPG technology for people older than 65. The AF prevalence increased from 8.5% for participants younger than 40 to 34.3% for those older than 65 (Figure 12). This increase in AF prevalence was expected because advanced age is a major risk factor for AF. The median number of measurements per participant increased along with the AF prevalence too (Figure 13). The more measurements a person performs, the higher the probability of gaining at least one true positive, false positive, or false negative. As the true negative measurements obtained the lowest priority, participants who performed a lot of measurements were more likely to become classified as true positive, false positive, or false negative on participant level. The specificity is hereby affected the most and its decrease should also lower the PPV in these groups. However, the increased AF prevalence in these older groups affected the PPV more positively and caused an overall increase along with the AF prevalence and an increased median number of measurements per participant.

Moreover, it was expected that the algorithm would classify measurements from healthy young adults with an increased heart rate variability (HRV) as false positives instead of true negatives and hereby lower the specificity of participants younger than 65. Although the specificity on participant level was at its lowest in the <40 age category with 74.8%, the 40-65 age group obtained the highest specificity of 83.3% (Figure 9). Therefore, there was no clear trend present that indicated the expected decrease of specificity along with a younger age. However, this may suggest that the more pronounced HRV that is present in adults younger than 40 does affect the specificity and requires more attention in future research.

Importance and future perspectives

This study provides novel insights into the algorithm's real-world performance and hereby demonstrates the impact that FibriCheck has on its users. Moreover, this work is an important addition to FibriCheck's transparency towards its (future) users by enlightening both the strengths and the

challenges that remain before FibriCheck can offer the algorithm as a standalone solution to detect AF. The findings of this study may also help FibriCheck's ongoing quest of investigating ways on how to improve the proprietary algorithm. In addition, this study emphasizes the important role of human revision in the current FibriCheck application. Therefore, it is suggested that training physicians in revising the PPG signals of their patients could greatly improve the accessibility of FibriCheck as a good candidate for AF detection and follow-up of patients who are diagnosed with AF. Both healthcare professionals and their patients benefit from the results of this study as it may increase the implementation of FibriCheck as a convenient, cost-effective, and non-invasive method for heart rhythm monitoring. Further research is required to determine if the algorithm's performance is influenced by the presence of certain risk factors. For this purpose, it may be interesting to study the algorithm's performance in different groups according to specific risk factors in a setting where there is a controlled measurement behavior. Although FibriCheck could benefit from a more specific algorithm, it would be unfavorable to diminish the highly sensitive nature to obtain a more specific one, as this would result in more false-negative results. Ideally, future versions of the algorithm will display both highly sensitive and highly specific behavior to function as a standalone solution for AF detection. The findings of this study suggest that FibriCheck can already benefit from the use of the proprietary algorithm as a standalone solution to rule out the presence of AF in the real-world population. However, this would require some adjustments in FibriCheck's workflow such as the implementation of a score that indicates just how sure the algorithm is about its findings. Evidently, the non-AF measurements would receive a very high score and would therefore not require human revision, while the possible AF measurements would receive a lower score and a subsequent human revision. This would allow for more efficient human revision and increase FibriCheck's cost-effectiveness and review capacity without loss of its accuracy.

Limitations

There are some limitations to this study that should be considered. Firstly, although the revision of all measurements was performed by experienced medical technicians under the supervision of a cardiologist, there was no 12-lead ECG available to confirm the diagnosis of AF according to the ESC recommendations. Secondly, as the results from the algorithm's automated analysis were visible to the reviewer, there might have been a certain bias by which the medical technician may have been more inclined to confirm the algorithm's findings, rather than correcting them. Thirdly, although one participant can perform multiple measurements, all measurements were considered to be independent for the statistical tests on measurement level. Finally, no follow-up data was available about the therapeutic interventions and diagnostic outcomes of participants with possible AF.

CONCLUSION

FibriCheck's AF detection algorithm showed promising results to rule out the presence of AF in the analyzed real-world population. However, further improvement of the proprietary algorithm is key to increase its implementation as a standalone solution for specific AF detection in the real-world setting. While FibriCheck is continuously investigating ways on how to improve the algorithm, it already manages to overcome this hurdle by offline human revisions that can be performed by either trained medical technicians or physicians. This synergy between the algorithm and human revision is crucial for the use of FibriCheck as a reliable tool for opportunistic AF screening and telemonitoring. In addition, this study suggests that the algorithm's performance varies across populations with differences in prevalence and measurement behavior. However, further research is required to determine if the presence of AF risk factors in such populations influence the algorithm's performance too.

REFERENCES

1. Mohan RA, Boukens BJ, Christoffels VM. Developmental Origin of the Cardiac Conduction System: Insight from Lineage Tracing. Pediatr Cardiol. 2018;39(6):1107-14.
2. Bakker ML, Moorman AF, Christoffels VM. The atrioventricular node: origin, development, and genetic program. Trends Cardiovasc Med. 2010;20(5):164-71.
3. Peters CH, Sharpe EJ, Proenza C. Cardiac Pacemaker Activity and Aging. Annu Rev Physiol. 2020;82:21-43.
4. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, et al. Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association. Circulation. 2019;139(10):e56-e528.
5. Krijthe BP, Kunst A, Benjamin EJ, Lip GY, Franco OH, Hofman A, et al. Projections on the number of individuals with atrial fibrillation in the European Union, from 2000 to 2060. Eur Heart J. 2013;34(35):2746-51.
6. Colilla S, Crow A, Petkun W, Singer DE, Simon T, Liu X. Estimates of current and future incidence and prevalence of atrial fibrillation in the U.S. adult population. Am J Cardiol. 2013;112(8):1142-7.
7. Staerk L, Sherer JA, Ko D, Benjamin EJ, Helm RH. Atrial Fibrillation: Epidemiology, Pathophysiology, and Clinical Outcomes. Circ Res. 2017;120(9):1501-17.
8. Allan V, Honarbakhsh S, Casas JP, Wallace J, Hunter R, Schilling R, et al. Are cardiovascular risk factors also associated with the incidence of atrial fibrillation? A systematic review and field synopsis of 23 factors in 32 population-based cohorts of 20 million participants. Thromb Haemost. 2017;117(5):837-50.
9. Heeringa J, van der Kuip DA, Hofman A, Kors JA, van Herpen G, Stricker BH, et al. Prevalence, incidence and lifetime risk of atrial fibrillation: the Rotterdam study. Eur Heart J. 2006;27(8):949-53.
10. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). Eur Heart J. 2021;42(5):373-498.
11. Wann LS, Curtis AB, Ellenbogen KA, Estes NA, Ezekowitz MD, Jackman WM, et al. Management of patients with atrial fibrillation (compilation of 2006 ACCF/AHA/ESC and 2011 ACCF/AHA/HRS recommendations): a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. Circulation. 2013;127(18):1916-26.
12. Davis EM, Packard KA, Knezevich JT, Campbell JA. New and emerging anticoagulant therapy for atrial fibrillation and acute coronary syndrome. Pharmacotherapy. 2011;31(10):975-1016.
13. Benjamin EJ, Wolf PA, D'Agostino RB, Silbershatz H, Kannel WB, Levy D. Impact of atrial fibrillation on the risk of death: the Framingham Heart Study. Circulation. 1998;98(10):946-52.
14. Jørgensen HS, Nakayama H, Reith J, Raaschou HO, Olsen TS. Acute stroke with atrial fibrillation. The Copenhagen Stroke Study. Stroke. 1996;27(10):1765-9.
15. Kimura K, Minematsu K, Yamaguchi T. Atrial fibrillation as a predictive factor for severe stroke and early death in 15,831 patients with acute ischaemic stroke. J Neurol Neurosurg Psychiatry. 2005;76(5):679-83.
16. Ceornodolea AD, Bal R, Severens JL. Epidemiology and Management of Atrial Fibrillation and Stroke: Review of Data from Four European Countries. Stroke Res Treat. 2017;2017:8593207.

17. Nattel S, Guasch E, Savelieva I, Cosio FG, Valverde I, Halperin JL, et al. Early management of atrial fibrillation to prevent cardiovascular complications. Eur Heart J. 2014;35(22):1448-56.
18. Van Gelder IC, Hemels ME. The progressive nature of atrial fibrillation: a rationale for early restoration and maintenance of sinus rhythm. Europace. 2006;8(11):943-9.
19. Goette A, Kalman JM, Aguinaga L, Akar J, Cabrera JA, Chen SA, et al. EHRA/HRS/APHRS/SOLAECE expert consensus on atrial cardiomyopathies: definition, characterization, and clinical implication. Europace. 2016;18(10):1455-90.
20. Gleason KT, Nazarian S, Dennison Himmelfarb CR. Atrial Fibrillation Symptoms and Sex, Race, and Psychological Distress: A Literature Review. J Cardiovasc Nurs. 2018;33(2):137-43.
21. Freedman B, Potpara TS, Lip GY. Stroke prevention in atrial fibrillation. Lancet. 2016;388(10046):806-17.
22. Martinez C, Katholing A, Freedman SB. Adverse prognosis of incidentally detected ambulatory atrial fibrillation. A cohort study. Thromb Haemost. 2014;112(2):276-86.
23. Nielsen JC, Lin YJ, de Oliveira Figueiredo MJ, Sepehri Shamloo A, Alfie A, Boveda S, et al. European Heart Rhythm Association (EHRA)/Heart Rhythm Society (HRS)/Asia Pacific Heart Rhythm Society (APHRS)/Latin American Heart Rhythm Society (LAHRS) expert consensus on risk assessment in cardiac arrhythmias: use the right tool for the right outcome, in the right population. J Arrhythm. 2020;36(4):553-607.
24. Klabunde RE. Cardiac electrophysiology: normal and ischemic ionic currents and the ECG. Adv Physiol Educ. 2017;41(1):29-37.
25. Taggar JS, Coleman T, Lewis S, Heneghan C, Jones M. Accuracy of methods for detecting an irregular pulse and suspected atrial fibrillation: A systematic review and meta-analysis. Eur J Prev Cardiol. 2016;23(12):1330-8.
26. Steinberg JS, O'Connell H, Li S, Ziegler PD. Thirty-Second Gold Standard Definition of Atrial Fibrillation and Its Relationship With Subsequent Arrhythmia Patterns: Analysis of a Large Prospective Device Database. Circ Arrhythm Electrophysiol. 2018;11(7):e006274.
27. Camm AJ, Kirchhof P, Lip GY, Schotten U, Savelieva I, Ernst S, et al. Guidelines for the management of atrial fibrillation: the Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC). Eur Heart J. 2010;31(19):2369-429.
28. Mairesse GH, Moran P, Van Gelder IC, Elsner C, Rosenqvist M, Mant J, et al. Screening for atrial fibrillation: a European Heart Rhythm Association (EHRA) consensus document endorsed by the Heart Rhythm Society (HRS), Asia Pacific Heart Rhythm Society (APHRS), and Sociedad Latinoamericana de Estimulación Cardíaca y Electrofisiología (SOLAECE). Europace. 2017;19(10):1589-623.
29. Ramkumar S, Nerlekar N, D'Souza D, Pol DJ, Kalman JM, Marwick TH. Atrial fibrillation detection using single lead portable electrocardiographic monitoring: a systematic review and meta-analysis. BMJ Open. 2018;8(9):e024178.
30. Steinhubl SR, Waalen J, Edwards AM, Ariniello LM, Mehta RR, Ebner GS, et al. Effect of a Home-Based Wearable Continuous ECG Monitoring Patch on Detection of Undiagnosed Atrial Fibrillation: The mSToPS Randomized Clinical Trial. Jama. 2018;320(2):146-55.
31. Proesmans T, Mortelmans C, Van Haelst R, Verbrugge F, Vandervoort P, Vaes B. Mobile Phone-Based Use of the Photoplethysmography Technique to Detect Atrial Fibrillation in Primary Care: Diagnostic Accuracy Study of the FibriCheck App. JMIR Mhealth Uhealth. 2019;7(3):e12284.

32. Elgendi M. On the analysis of fingertip photoplethysmogram signals. Curr Cardiol Rev. 2012;8(1):14-25.
33. Vandenberk T, Stans J, Van Schelvergem G, Pelckmans C, Smeets C, Lanssens D, et al. Clinical Validation of Heart Rate Apps: Mixed-Methods Evaluation Study. JMIR Mhealth Uhealth. 2017;5:e129.
34. Verbrugge FH, Proesmans T, Vijgen J, Mullens W, Rivero-Ayerza M, Van Herendael H, et al. Atrial fibrillation screening with photo-plethysmography through a smartphone camera. EP Europace. 2019;21(8):1167-75.
35. Selder JL, Proesmans T, Breukel L, Dur O, Gielen W, van Rossum AC, et al. Assessment of a standalone photoplethysmography (PPG) algorithm for detection of atrial fibrillation on wristband-derived data. Comput Methods Programs Biomed. 2020;197:105753.
36. Freedman B, Camm J, Calkins H, Healey JS, Rosenqvist M, Wang J, et al. Screening for Atrial Fibrillation: A Report of the AF-SCREEN International Collaboration. Circulation. 2017;135(19):1851-67.
37. Mortelmans C, Van Haelst R, Van Der Auwera J, Grieten L, Vandervoort P, Vaes B. 180Validation of a new smartphone application for the diagnosis of atrial fibrillation in primary care. EP Europace. 2017;19(suppl_3):iii16-iii.
38. Gruwez H, Evens S, Proesmans T, Smeets C, Haemers P, Pison L, et al. Evaluation of the device independent nature of a photoplethysmography-deriving smartphone app. EP Europace. 2021;23.
39. Apple I. Using Apple Watch for Arrhythmia Detection 2018.
40. Gruwez H, Evens S, Proesmans T, Duncker D, Linz D, Heidbuchel H, et al. The accuracy of physician interpretation of PPG vs single-lead ECG vs 12-lead ECG for the detection of atrial fibrillation. EP Europace. 2021;23(Supplement_3).
41. van der Velden RMJ, Verhaert DVM, Hermans ANL, Duncker D, Manninger M, Betz K, et al. The photoplethysmography dictionary: Practical guidance on signal interpretation and clinical scenarios from TeleCheck-AF. European Heart Journal - Digital Health. 2021.

Acknowledgements – The principal author of this paper is grateful for the opportunity to perform his senior internship in such a motivated and driven environment as FibriCheck. This internship was a unique and challenging experience that would not have been possible without the continuous support and close collaboration with the promotor Prof. dr. Lars Grieten and two daily supervisors MSc Stijn Evens and Ir. Glenn De Witte. All supervisors made an extensive effort to pro-actively support the principal author and managed to create a very nice and instructive teleworking experience.

Author contributions – LG, SE, and GDW conceived and designed this research project. JG performed the data analysis and wrote this paper with the support of LG, SE, and GDW. All authors were involved in the interpretation of the data.

SUPPLEMENTARY INFORMATION

Supplementary Table 1: Diagnostic metrics for the total population on measurement level and participant level.

	Measurement level (n = 409,299)	**Participant level (n = 2,163)**
True positives, n (%)	12,916	399
False positives, n (%)	6,140	346
True negatives, n (%)	390,069	1,414
False negatives, n (%)	174	4
Sensitivity, (%)	98.7	99.0
Specificity, (%)	98.5	80.3
Accuracy, (%)	98.5	83.8
PPV, (%)	67.8	53.6
NPV, (%)	100	99.7
Prevalence, (%)	3.2	18.6

Abbreviations: NPV, negative predictive value. PPV, positive predictive value.

Supplementary Table 2: Diagnostic metrics per age group on measurement level.

	< 40 years (n = 70,201)	**40-65 years (n = 220,159)**	**>65 years (n = 118,939)**
True positives, n	289	6,488	6,139
False positives, n	1,314	2,436	2,390
True negatives, n	68,591	211,126	110,352
False negatives, n	7	109	58
Sensitivity, (%)	97.6	98.4	99.1
Specificity, (%)	98.1	98.9	97.9
Accuracy, (%)	98.1	98.8	97.9
PPV, (%)	18.0	72.7	72.0
NPV, (%)	100	100	100
Prevalence, (%)	0.4	3	5.2
Measurements per participant, n, median (IQR)	39 (104.5)	56 (129)	85.5 (236.3)

Abbreviations: NPV, negative predictive value. PPV, positive predictive value.

Supplementary Table 3: Characteristics per age group. All 2,163 participants provided information about age and gender upon registration. Additional information was provided voluntarily by 2,085 **(**96.4%) participants via an in-app medical questionnaire.

	< 40 years (n = 495)	**40-65 years (n = 1,166)**	**>65 years (n = 502)**
Mandatory data points, n	**495**	**1,166**	**502**
Age, years, median (IQR)	33 (8)	54 (12)	71 (7)
Male, n (%)	231 (46.7)	614 (52.7)	343 (68.3)
Voluntary data points, n	**459**	**1,131**	**495**
Hypertension, n (%)	49 (10.7)	313 (27.7)	240 (48.5)
Diabetes mellitus, n (%)	14 (3.1)	72 (6.4)	65 (13.1)
Vascular disease, n (%)	9 (2)	70 (6.2)	80 (16.2)
Heart failure, n (%)	10 (2.2)	67 (5.9)	73 (14.8)
Stroke (CVA / TIA), n (%)	13 (2.8)	50 (4.4)	50 (10.1)
Pacemaker, n (%)	3 (0.7)	19 (1.7)	18 (3.6)
Arrhythmias, n (%)	127 (27.7)	479 (42.4)	343 (69.3)
Atrial fibrillation, n (%)	32 (7)	238 (21)	238 (48.1)
Blood thinner, n (%)	23 (5)	274 (24.2)	314 (63.4)
CHA_2DS_2-VASc score, median (IQR)	1 (1)	1 (1)	2 (1)

Abbreviations: IQR, interquartile range. AF, atrial fibrillation. CVA, cerebrovascular accident. TIA, transient ischemic attack. CHA_2DS_2-VASc, congestive heart failure, hypertension, age, diabetes mellitus, stroke, vascular disease, and sex category.

Supplementary Table 4: Diagnostic metrics per age group on participant level

	< 40 years (n = 495)	**40-65 years (n = 1,166)**	**> 65 years (n = 502)**
True positives, n	41	186	172
False positives, n	114	163	69
True negatives, n	339	814	261
False negatives, n	1	3	0
Sensitivity, (%)	97.6	98.4	100
Specificity, (%)	74.8	83.3	79.1
Accuracy, (%)	76.8	85.8	86.3
PPV, (%)	26.4	53.3	71.4
NPV, (%)	99.7	99.6	100
Prevalence, (%)	8.5	16.2	34.3

Abbreviations: NPV, negative predictive value. PPV, positive predictive value.

Supplementary Table 5: Diagnostic metrics per risk group on measurement level. The analysis per risk group was based on the 2,085 **(**96.4%) participants that voluntarily provided additional information via an in-app medical questionnaire. The remaining 78 (3.6%) participants did not provide the information necessary for CHA_2DS_2-VASc score calculation.

	Low risk (n = 160,038)	**Moderate risk (n = 104,299)**	**High risk (n = 128,115)**
True positives, n	2,422	3,179	6,644
False positives, n	2,685	1,117	2,259
True negatives, n	154,889	99,969	119,152
False negatives, n	42	34	60
Sensitivity, (%)	98.3	98.9	99.1
Specificity, (%)	98.3	98.9	98.1
Accuracy, (%)	98.3	98.9	98.2
PPV, (%)	47.5	74.1	74.6
NPV, (%)	100	100	100
Prevalence, (%)	1.5	3.1	5.2
Measurements per participant, n, median (IQR)	53 (121)	62 (160)	83 (214)

Abbreviations: NPV, negative predictive value. PPV, positive predictive value.

Supplementary Table 6: Characteristics per risk group. The analysis per risk group was based on the 2,085 **(**96.4%) participants that voluntarily provided additional information via an in-app medical questionnaire. The remaining 78 (3.6%) participants did not provide the information necessary for CHA_2DS_2-VASc score calculation.

	Low risk (n = 1,105)	**Moderate risk (n = 487)**	**High risk (n = 493)**
Mandatory data points, n	**1,105**	**487**	**493**
Age, years, median (IQR)	45 (18)	59 (17)	68 (12)
Male, n (%)	540 (49)	274 (56)	334 (68)
Voluntary data points, n	**1,105**	**487**	**493**
Hypertension, n (%)	0 (0)	270 (55.4)	332 (67.3)
Diabetes mellitus, n (%)	0 (0)	27 (5.5)	124 (25.2)
Vascular disease, n (%)	0 (0)	21 (4.3)	138 (28)
Heart failure, n (%)	0 (0)	36 (7.4)	114 (23.1)
Stroke (CVA / TIA), n (%)	0 (0)	0 (0)	113 (22.9)
Pacemaker, n (%)	7 (0.6)	15 (3.1)	18 (3.7)
Arrhythmias, n (%)	372 (33.7)	251 (51.5)	326 (66.1)
Atrial fibrillation, n (%)	141 (12.8)	143 (29.4)	224 (45.4)
Blood thinner, n (%)	86 (7.8)	174 (35.7)	351 (71.2)
CHA_2DS_2-VASc score, median (IQR)	1 (1)	1 (1)	3 (2)

Abbreviations: IQR, interquartile range. AF, atrial fibrillation. CVA, cerebrovascular accident. TIA, transient ischemic attack. CHA_2DS_2-VASc, congestive heart failure, hypertension, age, diabetes mellitus, stroke, vascular disease, and sex category.

Supplementary Table 7: Diagnostic metrics per risk group on participant level. The analysis per risk group was based on the 2,085 **(**96.4%) participants that voluntarily provided additional information via an in-app medical questionnaire. The remaining 78 (3.6%) participants did not provide the information necessary for CHA_2DS_2-VASc score calculation.

	Low risk (n = 1,105)	**Moderate risk (n = 487)**	**High risk (n = 493)**
True positives, n	134	100	157
False positives, n	204	76	57
True negatives, n	766	310	278
False negatives, n	1	1	1
Sensitivity, (%)	99.3	99.0	99.4
Specificity, (%)	79.0	80.3	83.0
Accuracy, (%)	81.5	84.2	88.2
PPV, (%)	39.7	56.8	73.4
NPV, (%)	99.9	99.7	99.6
Prevalence, (%)	12.2	20.7	32.1

Abbreviations: NPV, negative predictive value. PPV, positive predictive value.

Supplementary Table 8: Diagnostic metrics for participants with atrial fibrillation (AF) history and without AF history on measurement level. The analysis per AF history group was based on the 2,085 **(**96.4%) participants that voluntarily provided additional information via an in-app medical questionnaire. The remaining 78 (3.6%) of participants did not provide information about a possible history of AF.

	No history of AF (n = 243,660)	**History of AF (n = 148,792)**
True positives, n	3,255	8,990
False positives, n	3,601	2,460
True negatives, n	236,756	137,254
False negatives, n	48	88
Sensitivity, (%)	98.6	99.0
Specificity, (%)	98.5	98.2
Accuracy, (%)	98.5	98.3
PPV, (%)	47.5	78.5
NPV, (%)	100	99.9
Prevalence, (%)	1.4	6.1
Measurements per participant, n, median (IQR)	51 (125)	105 (211.3)

Abbreviations: NPV, negative predictive value. PPV, positive predictive value.

Supplementary Table 9: Characteristics of participants with atrial fibrillation (AF) history and without AF history. The analysis per AF history group was based on the 2,085 **(**96.4%) participants that voluntarily provided additional information via an in-app medical questionnaire. The remaining 78 (3.6%) participants did not provide information about a possible history of AF.

	No history of AF (n = 1,577)	**History of AF (n = 508)**
Mandatory data points, n	**1,577**	**508**
Age, years, median (IQR)	50.4 (21)	63.8 (15)
Male, n (%)	812 (51.5)	336 (66.1)
Voluntary data points, n	**1,577**	**508**
Hypertension, n (%)	391 (24.8)	211 (41.5)
Diabetes mellitus, n (%)	105 (6.7)	46 (9.1)
Vascular disease, n (%)	103 (6.5)	56 (11)
Heart failure, n (%)	61 (3.9)	89 (17.5)
Stroke (CVA / TIA), n (%)	68 (4.3)	45 (8.9)
Pacemaker, n (%)	18 (1.1)	22 (4.3)
Arrhythmias, n (%)	482 (30.6)	467 (91.9)
Atrial fibrillation, n (%)	0 (0)	508 (100)
Blood thinner, n (%)	269 (17.1)	342 (67.3)
CHA_2DS_2-VASc score, median (IQR)	1 (2)	2 (2)

Abbreviations: IQR, interquartile range. AF, atrial fibrillation. CVA, cerebrovascular accident. TIA, transient ischemic attack. CHA_2DS_2-VASc, congestive heart failure, hypertension, age, diabetes mellitus, stroke, vascular disease, and sex category.

Supplementary Table 10: Diagnostic metrics for participants with atrial fibrillation (AF) history and without AF history on participant level. The analysis per AF history group was based on the 2,085 **(**96.4%) participants that voluntarily provided additional information via an in-app medical questionnaire. The remaining 78 (3.6%) of participants did not provide information about a possible history of AF.

	No history of AF (n = 1,577)	**History of AF (n = 508)**
True positives, n	169	222
False positives, n	279	58
True negatives, n	1,129	225
False negatives, n	0	3
Sensitivity, (%)	100	98.7
Specificity, (%)	80.2	79.5
Accuracy, (%)	82.3	88.0
PPV, (%)	37.7	79.3
NPV, (%)	100	98.7
Prevalence, (%)	10.7	44.3

Abbreviations: NPV, negative predictive value. PPV, positive predictive value.