

Stworzona w Polsce aplikacja MoodMon dla pacjentów z zaburzeniami afektywnymi – wyniki badania skuteczności systemu opartego na algorytmach sztucznej inteligencji

MoodMon application for patients with affective disorders created in Poland – results of the study of the system's effectiveness based on artificial intelligence algorithms

Marlena Sokół-Szawłowska¹, Małgorzata Sochacka²

¹Psychiatryczna Poradnia Przykliniczna, Instytut Psychiatrii i Neurologii, Warszawa, Polska,

²Britenet Med sp. z o.o., Warszawa, Polska

Summary

Aim. In a group of patients diagnosed with BD and MDD, an analysis was conducted to evaluate the efficacy of AI algorithms in detecting mental state changes based on physical voice parameters

Material and methods. The MoodMon system was developed, including a mobile application for smartphones. In the first stage, the AI was trained using objective data and clinical assessments conducted by psychiatrists, which included 17-item versions of the HDRS (Hamilton Depression Rating Scale) and YMRS (Young Mania Rating Scale) scales, and the CGI (Clinical Global Impression) scale. The second stage was to further refine the AI using individual and population data and generate alerts when subtle changes in mental state were detected. Both stages of the study lasted a total of 944 days.

Results. Physical voice parameters can serve as biomarkers in affective disorders. The most effective in detecting changes in mental state were 19 specific physical voice parameters. The system showed high performance, with the following sensitivity (true positive rate – TPR) and specificity (true negative rate – TNR) values – for both diagnoses: TPR = 89.5%, TNR = 98.8%; BD: TPR = 89.6%, TNR = 98.9%; MDD: TPR = 89.1%, TNR = 98.5%. MoodMon helped to accurately monitor and predict changes in the mental state of patients with affective disorders.

Conclusions. The MoodMon system is an objective AI tool that effectively identifies the initial period of mental state changes in affective disorders based on physical voice parameters.

Słowa kluczowe: zaburzenia afektywne, sztuczna inteligencja, parametry fizyczne głosu

Key words: affective disorders, artificial intelligence, physical parameters of the voice

Wstęp

Globalnie rozpowszechnienie choroby afektywnej dwubiegunowej to około 2–3%, a depresji od 6,3 do 10,3%, co stanowi duże wyzwanie dla systemów ochrony zdrowia [1–4]. Przy czym pacjenci chętnie zgadzają się na monitorowanie swego stanu psychicznego za pomocą aplikacji [5–7]. Osoby cierpiące na zaburzenia psychiczne aktywnie korzystają z Internetu tak samo często jak populacja ogólna (80%), co sprzyja rozwojowi nowoczesnych rozwiązań w zakresie zdrowia psychicznego (e-zdrowie psychiczne) [8–12].

Rok 2010 wyznaczył początek rozwoju dziedziny badań nad wprowadzaniem rozwiązań technologicznych w zaburzeniach afektywnych [12–14]. Opracowano i przetestowano kilka systemów opartych na połączeniu dwóch typów danych: aktywnych-subiektywnych, tj. zgłaszanych przez samych użytkowników, i pasywnych-obiektywnych (behawioralnych lub fizjologicznych), tj. mierzonych przez czujniki osadzone w smartfonach lub w urządzeniach do nich podłączonych [12]. Ich najsilniejszym atutem jest wysoka przydatność do zbierania obiektywnych danych z monitorowania behawioralnego i fizjologicznego, takich jak: aktywność fizyczna, liczba rozmów telefonicznych oraz parametry fizyczne głosu [12, 15–18]. Potencjalnie mogłyby one ułatwiać ustalanie diagnozy stanów afektywnych i wczesną interwencję w wypadku objawów prodromalnych, a także umożliwić ciągłe monitorowanie [12, 19–23].

Oprócz zbierania danych istnieje już technologiczna możliwość testowania skuteczności algorytmów sztucznej inteligencji (Artificial Intelligence – AI) w przewidywaniu zmian stanu psychicznego w zaburzeniach afektywnych [12, 19, 24–30]. Polskim wkładem w rozwój wykorzystywania algorytmów AI w tym obszarze psychiatrii jest stworzenie systemu MoodMon.

Cel

W artykule opisano jeden z celów obszernego badania – ocenę skuteczności wysyłania alertów po wykryciu zmiany stanu psychicznego na podstawie fizycznych parametrów głosu w grupie pacjentów z zaburzeniami afektywnymi dwubiegunowymi i depresją nawracającą.

Zgodnie z naszą wiedzą przeprowadziłyśmy pierwsze na świecie badanie z udziałem tak licznej grupy, w którym system zbierał, analizował i wykorzystywał obiektywne dane głosowe pacjentów z chorobą afektywną dwubiegunową lub depresją nawracającą. Czas gromadzenia danych do analizy również był najdłuższy w porównaniu z innymi badaniami, bo przekroczył 12 miesięcy.

Material

Badanie trwało w latach 2021–2023 (zgoda Komisji Bioetycznej 170/2021). Zrekrutowano 100 pacjentów: 75 z chorobą afektywną dwubiegunową (typu I lub II) oraz 25 z zaburzeniami depresyjnymi nawracającymi. Uczestnicy udzielili świadomej zgody na udział w badaniu. Aby zapewnić dużą liczbę obserwowanych zmian, rekrutowano

tylko pacjentów ze stosunkowo szybką zmianą fazy. Kryteria włączenia wymagały co najmniej dwóch zmian fazy w ciągu ostatnich 12 miesięcy (+2 m-ce) u pacjentów z chorobą dwubiegunową i co najmniej jednego epizodu w tym okresie u pacjentów z depresją nawracającą.

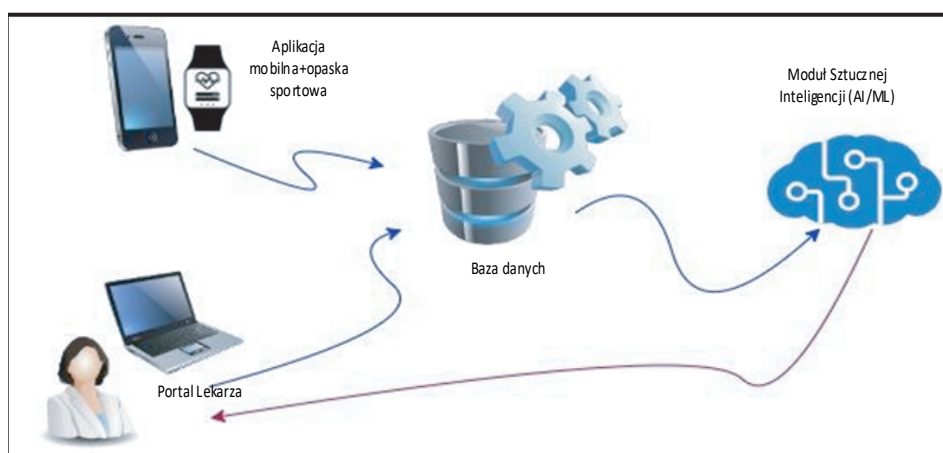
Po pierwszym (obserwacyjnym) etapie 12 pacjentów zrezygnowało z udziału w badaniu. Pozostało 88 badanych: 66 z chorobą afektywną dwubiegunową i 22 z zaburzeniami depresyjnymi nawracającymi. Przez ponad 18 miesięcy w badaniu uczestniczyło 84 pacjentów (ChAD-64; ChAJ-20). Co istotne, liczba pacjentów nie miała bezpośredniego wpływu na ilość i jakość zbieranych danych, gdyż zależały one od liczby zmian fazy, a zwłaszcza próbek głosu. Silnik analityczny (AI) trenowany był na danych zebranych od pacjentów w okresie trwania badania, który wynosił 944 dni (za zgodą Komisji Bioetycznej druga faza projektu została wydłużona).

Metody

System MoodMon stworzył zespół psychiatrów i specjalistów AI na bazie dostępnych badań. W skład systemu wchodzi:

- aplikacja mobilna (na smartfonach z systemem Android; zbiera i przesyła dane na bezpieczne serwery);
- krokomierz na nadgarstku (rejestruje aktywność fizyczną, parametry snu; zintegrowany z app);
- silnik analityczny oparty na AI (przetwarza zebrane dane i uczy się wysyłać alerty na podstawie obiektywnych danych z urządzeń i ocen klinicznych psychiatrów).

Schemat systemu MoodMon przedstawiono na rysunku 1.



Rysunek 1. Schemat funkcjonowania i składowych systemu MoodMon

Codzienna ankieta głosowa składała się z odpowiedzi na 3 pytania neutralne emocjonalnie. Na początku badania pacjent wybierał dogodną dla siebie porę nagrań (nie analizowano treści). Dokumentacja medyczna badania była przechowywana w eCRF.

Stan psychiczny oceniali ośmiu psychiatrów z dużym doświadczeniem w zaburzeniach afektywnych. Dane głosowe zbierane przez system MoodMon nie były dostępne dla badaczy i pacjentów. Dane z opaski monitorującej dotyczące aktywności ruchowej i snu były dostępne dla pacjentów oraz psychiatrów w postaci graficznej. W pierwszym etapie badania na wizycie wstępnej gromadzono dane kliniczne (w tym początkowy stan psychiczny) i demograficzne. Podczas kolejnych wizyt osobistych (planowo raz na 3 miesiące, a ze względu na pandemię COVID-19 dopuszczalny był format zdalny) stosowano podstawowe miary wyników: 17-punktowe wersje skal depresji HDRS (*Hamilton Depression Rating Scale*) i manii YMRS (*Young Mania Rating Scale*). Do ilościowego określenia oceny wykorzystano skalę CGI (*Clinical Global Impression*). Nasilenie objawów nie miało wpływu na kwalifikację do badania, jednak w wypadku pacjentów o bardzo dużym ich nasileniu (ocenionym na 6 lub 7 punktów w skali) konieczna była dokładna ocena kliniczna, czy zgoda na udział w badaniu została wyrażona świadomie. W tym badaniu wyniki HDRS i YMRS były uznawane jako wiążące w ocenie nasilenia objawów.

Podczas kolejnych wizyt pacjent był pytany o istotne zmiany w ogólnym stanie zdrowia i o inne ważne i/lub nagłe wydarzenia (ze względu na ich potencjalny wpływ na aktywność, nastrój, sen). Co dwa tygodnie psychiatra oceniał również stan pacjenta podczas krótkich/ screeningowych rozmów telefonicznych (z oceną CGI), aby ustalić, czy doszło do istotnej zmiany nastroju i aktywności od czasu poprzedniego kontaktu. Aby skrócić czas rozmowy telefonicznej, pacjent odpowiadał na pytania z wstępnie zdefiniowanego formularza wizyty. Pacjent był zapraszany na wizytę interwencyjną z pełną oceną z użyciem skal, gdy ankieta wskazywała, że taka zmiana wystąpiła.

Według zapisów z protokołu po badaniu psychiatra na podstawie doświadczenia klinicznego przypisywał 1 z 7 stanów spotykanych w przebiegu zaburzeń afektywnych, takich jak: mania, hipomania, zrównoważony nastrój (eutymia), subdepresja, depresja, ciężka depresja, stan mieszany. W systemie analizy danych powyższa ocena psychiatry była rejestrowana jako wartość CGI i stanowiła etykietę danych behawioralnych dostarczonych przez pacjenta w dniu oceny. W naszym badaniu celem było zidentyfikowanie wczesnych/najwcześniejszych i dyskretnych objawów zmiany fazy, dlatego przyjęliśmy niższe niż w poprzednich badaniach punkty odcięcia: eutymia – HDRS < 8 i YMRS < 6, depresja – HDRS ≥ 8 i YMRS < 6, hipomania/mania – HDRS < 8 i YMRS ≥ 6 oraz stan mieszany – HDRS ≥ 8 i YMRS ≥ 6 [18, 19, 30].

Dane behawioralne w sposób ciągły były zbierane przez smartfony, natomiast oceny kliniczne były znacznie rzadsze. Trening modelu predykcyjnego AI wymaga obecności obu typów danych, dlatego oceny psychiatryczne ekstrapolowano w okresach, w których stan psychiczny pacjentów nie był oceniany. Jeśli nie było zmiany CGI między dwiema kolejnymi wizytami, wszystkie dni w tym okresie były oznaczane CGI jako przypisane do dnia początkowego i końcowego. Jeśli zaś taką zmianę odnotowano, dane uzyskane między dwiema kolejnymi ocenami z różnymi CGI były wykluczane z zestawu danych do treningu.

W drugim etapie odbywały się wizyty w odstępach 3-miesięcznych oraz wizyty w reakcji na wysłany do lekarza przez system MoodMon alert. Po nim przeprowadzano krótkie wizyty telefoniczne (ocena CGI, pytania przesiewowe) lub gdy zachodziła konieczność kliniczna – dłuższe wizyty osobiste (ze skalami). Psychiatra oceniał zasadność alertu na podstawie zmiany lub braku zmiany stanu psychicznego. W ocenie uwzględniał również ważne okoliczności życiowe opisane przez pacjenta.

Dane były wstępnie przetwarzane, aby wyodrębnić te najbardziej istotne. Pod koniec pierwszego etapu wykazano, że fizyczne parametry głosu (m.in. wysokość, energia, współczynniki melcepralne) są najbardziej predykcyjne spośród zbieranych danych. Wybrano 19 o najwyższej wartości predykcyjnej i wykorzystano je do monitorowania nastroju pacjenta, aby informować poprzez alerty o nadchodzących zmianach w drugim etapie badania. Żeby przewidzieć możliwą zmianę nastroju i podjąć decyzję o wydaniu alertu, parametry akustyczne i pomocnicze (m.in. sezonowość, płeć) zostały przeanalizowane przez modele AI zarówno na poziomie populacji (dane od wszystkich pacjentów), jak i indywidualnym. Model był ponownie szkolony – służyły do tego wartości CGI określone przez psychiatrów podczas oceny alertu pod względem jego zasadności.

Szczególnie innowacyjnym aspektem był mechanizm wysyłania alertów włączony w drugiej fazie projektu. Psychiatra dzięki tej funkcji otrzymywał sygnał potencjalnej zmiany w stanie psychicznym pacjenta. W praktyce po powiadomieniu lekarz ma szansę przeprowadzić kompleksową ocenę, obejmującą nie tylko aktualny stan pacjenta, ale także trafność i pilność alertu. Najbardziej godną uwagi cechą tego procesu oceny jest nadrzędna warstwa ludzkiej wiedzy wobec zautomatyzowanego systemu. To lekarz na podstawie swojej wiedzy i doświadczenia klinicznego subiektywnie ocenia zasadność alertu i podejmuje decyzję o interwencji terapeutycznej bądź o jej zaniechaniu. W tym miejscu należy podkreślić, że system nie dostarcza informacji o kierunku zmiany stanu psychicznego w znaczeniu: „pogorszenie” czy „poprawa”. Dodatkowo wychwytuje i sygnalizuje zmiany stanu spowodowane innymi okolicznościami, które mogą nie być związane z chorobą (czasem to czynniki spustowe epizodów).

Po ocenie zasadności alertu system otrzymywał dane do obliczenia wskaźników „sukcesu” (określenie techniczne z AI, w znaczeniu „skuteczność”), zwłaszcza zaś do obliczenia prawdziwie dodatniego wskaźnika (*True Positive Rate* – TPR) i prawdziwie ujemnego wskaźnika (*True Negative Rate* – TNR). W ramce poniżej przedstawiono ogólną formę obliczeń. Bardziej szczegółowy opis uczenia maszynowego oraz rygorystycznego dochodzenia do najbardziej predykcyjnych akustycznych parametrów głosu przekracza ramy tego opracowania. Z tymi danymi można się zapoznać w technologicznym artykule autorek z naszego zespołu badawczego [10].

TPR (czułość – sensitivity, wskaźnik prawdziwie dodatnich wyników) – wskaźnik prawidłowo przewidzianych zmian nastroju

$$TPR = \frac{\Sigma tp}{\Sigma tp + \Sigma fn}$$

TNR (swoistość – specificity, wskaźnik prawdziwie negatywnych wyników) – wskaźnik prawidłowo przewidzianego braku zmian

$$TNR = \frac{\Sigma tn}{\Sigma tn + \Sigma fp}$$

Etykiety dla każdego dnia wszystkich pacjentów objętych obserwacją zostały oznaczone następująco:

tp – (true positive) prawdziwie pozytywny = alert wydany i oceniony jako uzasadniony (zmiana nastroju wystąpiła od czasu poprzedniej oceny klinicznej pacjenta),

tn – (true negative) prawdziwie negatywny = brak alertu i jego brak oceniony jako poprawny (brak zmiany nastroju od czasu poprzedniej oceny klinicznej pacjenta),

fp – (false positive) fałszywie pozytywny = alert wysłany, ale oceniony jako nieuzasadniony (brak zmiany nastroju od czasu poprzedniej oceny klinicznej pacjenta),

fn – (false negative) brak alertu i decyzja oceniona jako nieprawidłowa (zmiana nastroju wystąpiła od czasu poprzedniej oceny klinicznej pacjenta).

Wyniki

Średni czas trwania choroby badanych: ChAD – 8,6 lat; ChAJ – 8 lat. Inne dane demograficzne zamieszczono w tabeli 1.

Tabela 1. Dane demograficzne

Kategoria	ChAJ N = 25 (100%)	ChAD N = 75 (100%)
Przedział wieku		
18–24	6 (24%)	6 (8%)
25–34	6 (24%)	18 (24%)
35–44	6 (24%)	20 (26%)
45–54	6 (24%)	16 (22%)
55–64	1(4%)	15 (20%)
Płeć		
Mężczyzna	7(28%)	33 (44%)
Kobieta	18(72%)	42 (56%)

dalszy ciąg tabeli na następnej stronie

Stan cywilny		
Rozwiedziony/rozwidzona	2(8%)	11 (15%)
Stanu wolnego	13(52%)	35 (46%)
Żonaty/zamężna	10(40%)	27 (36%)
Wdowiec/wdowa	0(0%)	2 (3%)
Wykształcenie		
Wyższe	13(52%)	56 (75%)
Średnie	11(44%)	18 (24%)
Zawodowe	1(4%)	1 (1%)
Aktywność zawodowa		
Pracuje zawodowo	17(68%)	52 (69%)
Inne	3(12%)	4 (5,5%)
Rencista/emeryt	1(4%)	10 (13,5%)
Student lub uczeń	4(16%)	7 (9%)
Bezrobotny/bezrobotna	0(0%)	2 (3%)
Rodzaj zamieszkania		
Sam/sama	6(24%)	18 (24%)
Z partnerem/partnerką	3(12%)	13 (17%)
Z rodziną	16(64%)	44 (59%)
Wielkość miasta		
>200 000	15(60%)	57 (76%)
5000–200 000	5(20%)	13 (17,5%)
<5000	5(20%)	5 (6,5%)

W trakcie całego badania odnotowano 1394 zmiany stanu psychicznego pacjentów. Uzyskano 97 428 kompletów danych behawioralnych, z których 53 934 było pełnych i opatrzonych oceną kliniczną stanu pacjenta na dzień pobrania. W tym zaobserwowano 1344 zmiany, dla których były dostępne pełne dane i ocena kliniczna. Te zestawy danych służyły do trenowania modeli sztucznej inteligencji. W całym okresie obserwacji u 9 pacjentów nie odnotowano żadnej zmiany stanu psychicznego. Za pełny zestaw danych uznaje się również dane uzupełnione. Uzupełnianie danych odbywało się wtedy, gdy w zestawie dla danego dnia brakowało jednego lub kilku parametrów. Brakujące parametry uzupełniane były średnią z 14 poprzedzających dni, jeśli w tym

okresie nie brakowało danych z więcej niż 7 dni. Gdy taka średnia nie była możliwa do wyliczenia, cały zestaw był odrzucany.

W toku badania przeprowadzono 100 wizyt wstępnych, 410 osobistych, 1908 telefonicznych, 760 interwencyjnych i 100 końcowych. Odnotowano 2656 alarmów. Przez psychiatrów zostało ocenionych 1587 alarmów. Liczba faz odnotowanych dla pacjentów przy kolejnych kontaktach w toku dwóch etapów badania wyniosła 4589.

Efektywność systemu dla obu wskazań razem wyniosła: TPR = 89,5%, TNR = 98,8%; dla zaburzeń afektywnych dwubiegunowych: TPR = 89,6%, TNR = 98,9%; dla zaburzeń depresyjnych nawracających: TPR = 89,1%, TNR = 98,5%.

Omówienie wyników

Badania kliniczne analizujące zmiany w języku mówionym w zaburzeniach afektywnych w zależności od etapu choroby pochodzą z 1938 roku [31]. Bez dostępnych obecnie zaawansowanych technologii badano również osoby zdrowe. W czasie tych analiz dowiedziono, że konkretny afekt wokalny koreluje z określonymi wzorcami modulacji głosu. Podczas przeżywania różnych emocji dochodzi do modyfikacji wzorców oddychania, fonacji lub artykulacji – znajduje to odzwierciedlenie w sygnale mowy. Nawet cechy podstawowe wypowiedzi, takie jak tempo, głośność, intonacja czy rytm, mogą różnicować stany afektywne. Przeżywanie gniewu oraz lęku prowadzi do przyspieszenia tempa z wysokimi wartościami częstotliwości podstawowej (F_0) i szerokim zakresem intonacji z powodu aktywacji układu współczulnego, szybkim tętnem i zwiększonym ciśnieniem krwi, czemu czasami może towarzyszyć suchość w ustach i drżenie mięśni. Odwrotnie dzieje się w sytuacji nudy czy przeżywania smutku [32, 33]. Gdy mowa jest wolna i monotonna, F_0 jest obniżane bez większych zmian w intonacji. Jest to spowodowane stymulacją układu nerwowego przywspółczulnego, spowolnieniem akcji serca, spadkiem ciśnienia krwi i zwiększoną produkcją śliny. Jednak bardziej złożone wzorce także okazały się obiecujące w rozpoznawaniu stanów emocjonalnych. Należą do nich liniowe parametry kodowania predykcyjnego (*Linear Predictive Coding Parameters* – LPC) lub współczynniki częstotliwości cepstrum Mel (*Mel-frequency Cepstrum Coefficients* – MFCC) [34].

Postęp technologiczny ostatnich lat przyczynił się do zgłębienia zagadnienia zmienności fizycznych parametrów głosu w okresach remisji i trwania epizodów zaburzeń afektywnych. W zdecydowanej większości były to dotychczas krótkie badania na nielicznych lub bardzo nielicznych populacjach klinicznych, w których była stosowana zróżnicowana metodologia.

W badaniu pilotażowym Karama i wsp. z 2014 roku zbierano dane dotyczące mowy w okresie 6–12 miesięcy od 6 uczestników (4 kobiet i 2 mężczyzn) z diagnozą choroby afektywnej dwubiegunowej typu I i historią szybkich zmian nastroju (czterema lub więcej epizodami manii, hipomanii lub depresji rocznie). Próbkę głosu pobierano w czasie nagrań w toku interakcji klinicznych i interakcji nieklinicznych (dokonano opisu metodologii do ciągłego i dyskretnego zbierania niestrukturyzowanej mowy poprzez nagrywanie codziennych rozmów telefonicznych). We wnioskach wskazano, że stany maniakalne i depresyjne można rozpoznać na podstawie fizycznych danych

dotyczących mowy (część z nich była bardziej predykcyjna). Zaznaczono jednak, że system miał trudności z wykrywaniem depresji na podstawie nagrań głosu podczas interakcji nieklinicznych. W tym badaniu najbardziej informatywnymi cechami dla klasyfikacji stanów dwubiegunowych okazały się: średnia wykrywana binarna aktywność dźwięczna, *SD* wysokości dźwięku, średnia segmentowa szybkości przejścia przez zero i średnia segmentowa wygładzonej aktywności dźwięcznej [35].

Muaremi i wsp. w 2014 roku opisali wykonalność analizy głosu podczas rozmowy telefonicznej w celu przewidywania obecności epizodów choroby dwubiegunowej. Uwzględniono naturalne rozmowy telefoniczne 12 pacjentów z 12 tygodni. Dodatkowo były przeprowadzane wywiady z psychiatrą co 3 tygodnie w szpitalu. Wykazano, że najważniejszymi cechami mowy wykorzystywanymi do przewidywania stanów dwubiegunowych były: wartość stosunku harmoniczných do szumu (HNR), liczba krótkich obrotów i wariancja wysokości dźwięku *F0* [36].

W 2015 roku Guidi i wsp. opisali studium przypadku (dane od 1 mężczyzny w wieku 36 lat z chorobą afektywną dwubiegunową, używającego smartfonu z systemem Android – najbardziej rozpowszechnionym wśród użytkowników telefonów na świecie). W tym badaniu zaprojektowano aplikację do analizy bieżącej mowy za pomocą urządzenia typu smartfon. Porównano 30-sekundowe próbki głosu pacjenta z próbkami od 1 kobiety (25 lat) i 1 mężczyzny (30 lat) bez zaburzeń psychicznych. Badanie trwało 14 tygodni. Zadanie dla uczestników polegało na komentowaniu obrazu (wykonano je 15 razy w domu). Nastroj pacjenta był oceniany przez psychiatrę w przeddzień każdej sesji nagrywania głosu. Aplikacja mogła rejestrować próbki audio i szacować podstawową częstotliwość mowy *F0* i jej zmiany. Wywnioskowano, że zmiany w cechach wysokości głosu można wyodrębnić za pomocą smartfonów i że korelują one znacząco ze zmianami między stanem wyrównanego nastroju, depresji i hipomaniakalnym u pacjenta z chorobą afektywną dwubiegunową [37].

W badaniu Gideona i wsp. opisanym w 2016 roku monitorowano wzorce mowy w życiu rzeczywistym. Założono, że w warunkach naturalistycznych pacjenci używają różnych modeli smartfonów, co powoduje zmiany akustyczne podczas nagrań głosu. Przeanalizowano dane od 37 uczestników chorujących na chorobę dwubiegunową w średnim okresie 29 tygodni. Odnotowano 780 cotygodniowych ocen klinicznych. Do analizy wykorzystano tylko te ustrukturyzowane połączenia. W artykule przedstawiono metody poprawy porównywalności zebranych danych, wykorzystując różne podejścia do wstępnego przetwarzania, ekstrakcji cech i modelowania danych. Autorzy byli w stanie zwiększyć moc dyskryminacyjną swoich algorytmów przewidywania nastroju, w którym aktualnie był pacjent. Zaznaczono, że poważnym ograniczeniem badania było analizowanie tylko ustrukturyzowanych rozmów kwalifikacyjnych z różnych urządzeń [38].

Badanie Maxhuniego i wsp. opierało się na założeniu, że aktywność motoryczna związana z mową jest jednym z rodzajów aktywności motorycznej, a ta stanowi najbardziej spójny wskaźnik choroby afektywnej dwubiegunowej. Dane audio, dane z akcelerometru i samooceny pochodziły od 5 pacjentów z chorobą afektywną dwubiegunową. Analizowano okres 12 tygodni podczas rzeczywistych aktywności. Przedstawiono system zdolny do klasyfikowania stanu pacjentów cierpiących na tę chorobę

z użyciem informacji sensorycznych ze smartfonów. Żadne wcześniejsze badania nie skupiały się na naturalnej obserwacji codziennych rozmów telefonicznych w celu klasyfikowania upośledzonego funkcjonowania życiowego u osób z chorobą afektywną dwubiegunową. Uzyskano wyniki wskazujące na możliwość klasyfikowania z dużą pewnością przebiegu epizodów nastroju lub nawrotów u pacjentów ze zdiagnozowaną tą jednostką. Wykorzystanie cech audio (emocjonalnych i widmowych) skutkowało wynikami o dokładności >80% w klasyfikacji nastroju pacjentów z chorobą afektywną dwubiegunową. Testowanie cech emocjonalnych i widmowych oddzielnie i łącznie dawało podobną dokładność [39].

Faurholt-Jepsen i wsp. w 2016 roku podali dokładność w zakresie od 0,61 do 0,74 w klasyfikacji nastroju na podstawie fizycznych cechy głosu. Wyniki te uzyskano w toku analizy danych pochodzących od 28 pacjentów ambulatoryjnych z chorobą afektywną dwubiegunową, gromadzonych w naturalnych warunkach codziennie przez okres 12 tygodni. Użyto cech głosu (wyodrębnionych podczas codziennych rozmów telefonicznych), automatycznie generowanych obiektywnych danych ze smartfona na temat aktywności behawioralnej oraz elektronicznych danych samokontrolnych. Cechy głosu okazały się dokładniejsze, bardziej czułe i specyficzne w klasyfikacji stanów maniakalnych lub mieszanych niż depresyjnych. Połączenie cech głosu z automatycznie generowanymi obiektywnymi danymi ze smartfona na temat aktywności behawioralnych i danymi samokontrolowanymi elektronicznie nieznacznie zwiększyło dokładność, czułość i specyficzność klasyfikacji stanów afektywnych. Dlatego wywnioskowano, że cechy głosu zebrane w naturalnych warunkach za pomocą smartfonów mogą być wykorzystywane jako obiektywne markery stanu u pacjentów z chorobą afektywną dwubiegunową [19].

Z analiz przytoczonych powyżej oraz dodatkowo z podsumowania przeglądu kilku innych (metodologicznie zróżnicowanych) badań wynika, że istnieje wiele fizycznych cech mowy, które można wytypować jako potencjalne biomarkery stanów afektywnych w depresji i chorobie dwubiegunowej. Funkcjonuje już wstępny konsensus (na bazie dostępnych fragmentarycznych badań) co do tego, że zmiany aktywności mowy i cech głosu mogą być potencjalnymi czułymi i trafnymi miarami obiektywnych objawów prodromalnych w epizodach afektywnych. A dzięki dalszemu wykorzystywaniu technologii AI w tym obszarze pojawiają się szanse na zastosowanie kliniczne (pomocne i wyczekiwane zarówno przez pacjentów, jak i psychiatrów) [12, 19, 27].

Do rozpoznawania depresji na podstawie cech akustycznych mowy najczęściej typowane były: prozodie (takie jak wysokość dźwięku, częstotliwość podstawowa (F_0), energia, tempo mówienia i pauzy), cechy jakości głosu (np. drżenie i migotanie), deskryptory traktu głosowego (formanty) oraz cechy widmowe (takie jak MFCC i LPC) [40–49]. Niedawno opublikowano najbardziej zaawansowane technologicznie wyniki badań nad fizycznymi parametrami głosu w depresji. Korzystają one – tak jak nasze badanie – z nowoczesnych możliwości analitycznych opartych na AI [50–52].

Na podstawie zdecydowanie jeszcze mniej licznych badań zidentyfikowano kilka potencjalnych biomarkerów manii w postaci fizycznych cech mowy. Zaliczono do nich prozodie, takie jak wysokość dźwięku, intensywność (energia), tempo mowy i czas trwania pauzy, LPC i widmowe współczynniki MFCC [53, 54].

Zatem podsumowując, można stwierdzić, że przed systemem MoodMon stworzono i testowano kilka podobnych rozwiązań, jednak o innych funkcjonalnościach. Opierano się na danych z krótkich okresów uzyskanych od niewielkich populacji.

Innowacyjność systemu MoodMon sprawia, że trudno go porównywać z innymi rozwiązaniami, gdyż przytoczone powyżej badania gromadziły obiektywne dane „statyczne” świadczące o remisji lub epizodzie, bez komponentu uczenia maszynowego w celu przewidywania zmian stanu psychicznego. W naszym projekcie AI wykorzystywano do przewidywania i „podpowiadania” klinicyście dyskretnych zmian stanu psychicznego sugerujących (za pomocą alertów) potrzebę kontaktu z pacjentem w celu oceny klinicznej na wypadek początku epizodu afektywnego. Po ocenie klinicznej w toku uczenia maszynowego system doskonalił się i zyskiwał na przyszłość coraz wyższą trafność w rozróżnianiu zmian stanu psychicznego wynikających z nawrotu depresji lub epizodu choroby afektywnej dwubiegunowej.

Dzięki niemu zgromadzono unikatowo duży zasób danych w postaci fizycznych parametrów głosu pochodzących od licznej populacji. Badanie wyróżnia się testowaniem AI jednocześnie w populacji pacjentów cierpiących na zaburzenia depresyjne nawracające oraz chorobę afektywną dwubiegunową [12, 27, 50–54]. Przeanalizowano największą ilość danych behawioralnych i danych pochodzących z wizyt, określonych jako pełny zestaw danych [12, 19, 27]. W toku badania uchwyciono aż 4570 faz i przypisano im etykiety jednego z siedmiu stanów afektywnych.

W czasie badania algorytmy AI łączyły liczne obiektywne parametry z oceną kliniczną. Analizy statystyczne wykazały, że w zestawie obiektywnych danych zbieranych przez system MoodMon z urządzeń mobilnych w badanej populacji najsilniej ze stanem psychicznym korelowało 19 fizycznych parametrów głosu [10]. Potwierdzono zatem, że zmiany fizycznych cech głosu odzwierciedlają zmiany w stanie psychicznym i mogą być traktowane jako biomarkery.

Lekarz psychiatra na bieżąco oceniał, czy AI trafnie wykonuje powierzone jej zadanie wysyłania alertu przy zmianie stanu psychicznego – oceny były tu najliczniejsze w porównaniu z innymi badaniami. Ten fakt przekładał się na doskonalenie trafności AI. System analizy działał zarówno na danych indywidualnych, jak i grupowych, toteż nasze ilości danych służących do trenowania były większe niż we wcześniejszych badaniach [12, 19, 27]. Wyniki o wysokiej czułości i swoistości alertów dla obu zaburzeń świadczą o tym, że MoodMon ma duży potencjał kliniczny.

Bardziej szczegółowe opisy różnicujące badane grupy pod względem m.in. czynników socjodemograficznych czy subiektywnego odbioru aplikacji MoodMon przez użytkowników w zależności od rodzaju zaburzenia afektywnego przekraczają ramy tego opracowania, dlatego zostały przedstawione w osobnych publikacjach [55, 56]. Odrębna technologiczna publikacja z niezbędnymi technicznymi wyjaśnieniami dla specjalistów AI zawiera detale dotyczące najbardziej predykcyjnego zestawu 19 fizycznych cech głosu pacjentów [10].

Wnioski

1. Nasze badanie z użyciem AI potwierdziło, że fizyczne parametry głosu są biomarkerami w zaburzeniach afektywnych.
2. Wysoka czułość i swoistość alertów MoodMon sprawia, że system może być stosowany w bardzo wczesnej fazie zmiany stanu psychicznego.
3. Nieodzowny i nadrzędny jest nadal czynnik ludzki – praca psychiatry polega na ocenie zasadności klinicznej wysyłania alertu przez AI i na podejmowaniu na tej podstawie decyzji terapeutycznych bądź ich zaniechaniu.

Piśmiennictwo

1. Nierenberg AA, Agustini B, Köhler-Forsberg O, Cusin C, Katz D, Sylvia LG i wsp. *Diagnosis and treatment of bipolar disorder: A review*. JAMA 2023; 330(14): 1370–1380.
2. Grande I, Berk M, Birmaher B, Vieta E. *Bipolar disorder*. Lancet 2016; 387(10027): 1561–1572.
3. Baumeister H, Härter M. *Prevalence of mental disorders based on general population surveys*. Soc. Psychiatry Psychiatr. Epidemiol. 2007; 42(7): 537–546.
4. Hayes JF, Miles J, Walters K, King M, Osborn DP. *A systematic review and meta-analysis of premature mortality in bipolar affective disorder*. Acta Psychiatr. Scand. 2015; 131(6): 417–425.
5. Daus H, Kislicyn N, Heuer S, Backenstrass M. *Disease management apps and technical assistance systems for bipolar disorder: Investigating the patients' point of view*. J. Affect. Disord. 2018; 229: 351–357.
6. Van Til K, McInnis MG, Cochran A. *A comparative study of engagement in mobile and wearable health monitoring for bipolar disorder*. Bipolar Disord. 2020; 22(2): 182–190. Doi: 10.1111/bdi.12849.
7. Dickerson RF, Gorlin EI, Stankovic JA. *Empath: A continuous remote emotional health monitoring system for depressive illness*. 2011. <https://www.cs.virginia.edu/~stankovic/psfiles/robempath.pdf>.
8. Trefflich F, Kalckreuth S, Mergl R, Rummel-Kluge C. *Psychiatric patients' internet use corresponds to the Internet use of the general public*. Psychiatry Res. 2015; 226(1): 136–141.
9. Jędrzejewska B, Borycka A, Kotulska M, Laskus P, Lichman M, Lubczyńska Z i wsp. *Efficiency of mobile apps for monitoring and management of mental health – Review article*. J. Educ. Health Sport 2021; 35(1): 62–80.
10. Kamińska D, Kamińska O, Sochacka M, Sokół-Szawłowska M. *The role of selected speech signal characteristics in discriminating unipolar and bipolar disorders*. Sensors 2024; 24(14): 4721. Doi: 10.3390/s24144721.
11. Hidalgo-Mazzei D, Mateu A, Reinares M, Matic A, Vieta E, Colom F. *Internet-based psychological interventions for bipolar disorder: Review of the present and insights into the future*. J. Affect. Disord. 2015; 188: 1–13.
12. Dogan E, Sander C, Wagner X, Hegerl U, Kohls E. *Smartphone-based monitoring of objective and subjective data in affective disorders: Where are we and where are we going? Systematic review*. J. Med. Internet Res. 2017; 19(7): e262.
13. Høifødt RS, Lillevoll KR, Griffiths KM, Wilsgaard T, Eisemann M, Waterloo K i wsp. *The clinical effectiveness of web-based cognitive behavioral therapy with face-to-face therapist*

- support for depressed primary care patients: Randomized controlled trial. *J. Med. Internet Res.* 2013; 15(8): e153.
14. Festersen P, Corradini A. Re: *Mind: A mobile application for bipolar disorder patients*. W: *4th International Conference on Wireless Mobile Communication and Healthcare – Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH), December 5, 2014, Athens, Greece*. 2014. S. 343–346.
 15. Crisan C. *Lack of insight in bipolar disorder: The impact on treatment adherence, adverse clinical outcomes and quality of life*. W: *Psychotic disorders – An update*. Budapest: IntechOpen; 2018.
 16. Faurholt-Jepsen M, Vinberg M, Frost M, Christensen EM, Bardram JE, Kessing LV. *Smartphone data as an electronic biomarker of illness activity in bipolar disorder*. *Bipolar Disord.* 2015; 17(7): 715–728.
 17. Grünerbl A, Muaremi A, Osmani V, Bahle G, Ohler S, Tröster G i wsp. *Using smartphone mobility traces for the diagnosis of depressive and manic episodes in bipolar patients*. W: *Proceedings of the 5th Augmented Human International Conference, March, 2014, Kobe, Japan*. 2014. S. 1–8.
 18. Grünerbl A, Muaremi A, Osmani V, Bahle G, Ohler S, Tröster G i wsp. *Smartphone-based recognition of states and state changes in bipolar disorder patients*. *IEEE J. Biomed. Health Inform.* 2015; 19(1): 140–148.
 19. Faurholt-Jepsen M, Busk J, Frost M, Vinberg M, Christensen EM, Winther O i wsp. *Voice analysis as an objective state marker in bipolar disorder*. *Transl. Psychiatry* 2016; 6(7): e856.
 20. Vieta E, Salagre E, Grande I, Carvalho AF, Fernandes BS, Berk M i wsp. *Early intervention in bipolar disorder*. *Am. J. Psychiatry* 2018; 175(5): 411–426.
 21. Altamura AC, Buoli M, Caldiroli A, Caron L, Cumerlato Melter C, Dobrea C i wsp. *Misdiagnosis, duration of untreated illness (DUI) and outcome in bipolar patients with psychotic symptoms: A naturalistic study*. *J. Affect. Disord.* 2015; 182: 70–75. Doi: 10.1016/j.jad.2015.04.024.
 22. Menculini G, Verdolini N, Gobbicchi C, Del Bello V, Serra R, Brustenghi F i wsp. *Clinical and psychopathological correlates of duration of untreated illness (DUI) in affective spectrum disorders*. *Eur. Neuropsychopharmacol.* 2022; 61: 60–70.
 23. Francis A, Gasparo P. *Interval between symptom onset and hospitalization in mania*. *J. Affect. Disord.* 1994; 31(3): 179–185.
 24. Faurholt-Jepsen M, Vinberg M, Christensen EM, Frost M, Bardram J, Kessing LV. *Daily electronic self-monitoring of subjective and objective symptoms in bipolar disorder – The MONARCA trial protocol (MONitoring, treAtment and pRediCtion of bipolar disorder episodes): A randomised controlled single-blind trial*. *BMJ Open* 2013; 3(7): e003353.
 25. Gentili C, Valenza G, Nardelli M, Lanatà A, Bertschy G, Weiner L i wsp. *Longitudinal monitoring of heartbeat dynamics predicts mood changes in bipolar patients: A pilot study*. *J. Affect. Disord.* 2017; 209: 30–38.
 26. Tams S, Hill K, Ortiz de Guinea A, Thatcher J, Grover V. *NeuroIS-alternative or complement to existing methods? Illustrating the holistic effects of neuroscience and self-reported data in the context of technostress research*. *J. Assoc. Inf. Syst.* 2014; 15(10): 723–753.
 27. Flanagan O, Chan A, Roop P, Sundram F. *Using acoustic speech patterns from smartphones to investigate mood disorders: Scoping review*. *JMIR Mhealth Uhealth.* 2021; 9(9): e24352.
 28. Dey A, Wac K, Ferreira D, Tassini K, Hong J, Ramos J. *Getting closer: An empirical investigation of the proximity of user to their smart phones*. W: *Proceedings of the 13th International Conference on Ubiquitous Computing*. New York: USA: ACM; 2011. S. A.

29. Gravenhorst F, Muaremi A, Bardram J, Grünerbl A, Mayora O, Wurzer G i wsp. *Mobile phones as medical devices in mental disorder treatment: An overview*. Pers. Ubiquit. Comput. 2014; 19(2): 335–353.
30. Faurholt-Jepsen M, Vinberg M, Frost M, Debel S, Christensen EM, Bardram JE i wsp. *Behavioral activities collected through smartphones and the association with illness activity in bipolar disorder*. Int. J. Methods Psychiatr. Res. 2016; 25(4): 309–323.
31. Newman S, Mather VG. *Analysis of spoken language of patients with affective disorders*. Am. J. Psychiatry 1938; 94: 913–942.
32. Picard RW. *Affective computing*. Cambridge, MA: MIT Press; 2000.
33. Williams CE, Stevens KN. *Emotions and speech: Some acoustical correlates*. J. Acoust. Soc. Am. 1972; 52(4): 1238–1250.
34. Nakatsu R, Nicholson J, Tosa N. *Emotion recognition and its application to computer agents with spontaneous interactive capabilities*. W: *Proceedings of the 7th ACM International Conference on Multimedia (Part 1), 1999 Oct 30–Nov 5, Orlando, FL*. New York: ACM; 1999. S. 343–351.
35. Karam ZN, Provost EM, Singh S, Montgomery J, Archer C, Harrington G i wsp. *Ecologically valid long-term mood monitoring of individuals with bipolar disorder using speech*. Proc. IEEE Int. Conf. Acoust. Speech Signal Process. 2014; 2014: 4858–4862.
36. Muaremi A, Gravenhorst F, Grünerbl A, Arnrich B, Tröster G. *Assessing bipolar episodes using speech cues derived from phone calls*. W: *Cipresso P, Matic A, Lopez G. red. Pervasive computing paradigms for mental health. MindCare 2014*. Cham: Springer; 2014. S. 103–114.
37. Guidi A, Salvi S, Ottaviano M, Gentili C, Bertschy G, de RD i wsp. *Smartphone application for the analysis of prosodic features in running speech with a focus on bipolar disorders: System performance evaluation and case study*. Sensors (Basel) 2015; 15(11): 28070–2887.
38. Gideon J, Provost EM, McInnis M. *Mood state prediction from speech of varying acoustic quality for individuals with bipolar disorder*. Proc IEEE Int. Conf. Acoust. Speech Signal Process 2016; 2016: 2359–2363.
39. Maxhuni A, Muñoz-Meléndez A, Osmani V, Perez H, Mayora O, Morales EF. *Classification of bipolar disorder episodes based on analysis of voice and motor activity of patients*. Pervasive Mob. Comput. 2016; 31: 50–66.
40. Cohn JF, Kruez TS, Matthews I, Yang Y, Nguyen MH, Padilla MT i wsp. *Detecting depression from facial actions and vocal prosody*. W: *Proceedings of the 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, 2009 Sep 10–12, Amsterdam, Netherlands*. IEEE; 2009. S. 1–7.
41. Yang Y, Fairbairn C, Cohn JF. *Detecting depression severity from vocal prosody*. IEEE Trans. Affect. Comput. 2013; 4(2): 142–150.
42. Alghowinem S, Goecke R, Wagner M, Epps J, Breakspear M, Parker G. *Detecting depression: A comparison between spontaneous and read speech*. W: *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing; 2013 May 26–31, Vancouver, BC, Canada*. IEEE; 2013. S. 7547–7551.
43. Cummins N, Vlasenko B, Sagha H, Schuller B. *Enhancing speech-based depression detection through gender dependent vowel-level formant features*. W: *Proceedings of the Conference on Artificial Intelligence in Medicine in Europe, 2017 Jun 21–24, Vienna, Austria*. Springer; 2017. S. 209–214.
44. Moore E, Clements M, Peifer J, Weisser L. *Comparing objective feature statistics of speech for classifying clinical depression*. W: *Proceedings of the 26th Annual International Conference*

- of the IEEE Engineering in Medicine and Biology Society, 2004 Sep 1–5, San Francisco, CA.* IEEE; 2004. S. 17–20.
45. Rejaibi E, Komaty A, Meriaudeau F, Agrebi S, Othmani A. *MFCC-based recurrent neural network for automatic clinical depression recognition and assessment from speech.* Biomed. Signal Process. Control. 2022; 71(Part A): 103107.
 46. Cummins N, Epps J, Breakspear M, Goecke R. *An investigation of depressed speech detection: Features and normalization.* W: *Proceedings of the 12th Annual Conference of the International Speech Communication Association, 2011 Aug 27–31, Florence, Italy.* Interspeech; 2011.
 47. Taguchi T, Tachikawa H, Nemoto K, Suzuki M, Nagano T, Tachibana R i wsp. *Major depressive disorder discrimination using vocal acoustic features.* J. Affect. Disord. 2018; 225: 214–220.
 48. Lopez-Otero P, Docio-Fernandez L, Garcia-Mateo C. *Assessing speaker independence on a speech-based depression level estimation system.* Pattern Recognit. Lett. 2015; 68(Part 2): 343–350.
 49. Ringeval F, Schuller B, Valstar M, Gratch J, Cowie R, Scherer S i wsp. *AVEC 2017: Real-life depression and affect recognition workshop and challenge.* W: *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge, 2017 Oct 23, Mountain View, CA.* ACM; 2017. S. 3–9.
 50. Weisenburger RL, Mullarkey MC, Labrada J, Labrousse D, Yang MY, MacPherson AH i wsp. *Conversational assessment using artificial intelligence is as clinically useful as depression scales and preferred by users.* J. Affect. Disord. 2024; 351: 489–498.
 51. Zhang Y, Folarin AA, Dineley J, Conde P, Angel de V, Sun S i wsp. *Identifying depression-related topics in smartphone-collected free-response speech recordings using an automatic speech recognition system and a deep learning topic model.* J. Affect. Disord. 2024; 355: 40–49.
 52. Ravi V, Wang J, Flint J, Alwan A. *Enhancing accuracy and privacy in speech-based depression detection through speaker disentanglement.* Comput. Speech Lang. 2024; 86: 101605.
 53. Farrús M, Codina-Filbà J, Escudero J. *Acoustic and prosodic information for home monitoring of bipolar disorder.* Health Informatics J. 2021; 27(1): 1460458220972755.
 54. Bedi G, Carrillo F, Cecchi GA, Slezak DF, Sigman M, Mota NB i wsp. *Automated analysis of free speech predicts psychosis onset in high-risk youths.* NPJ Schizophr. 2015; 1: 15030.
 55. Sokół-Szawłowska M, Kamińska O, Sochacka M. *MoodMon: novel optimization of bipolar disorder monitoring through patient-driven voice parameter submission and AI technology.* Advances in Psychiatry and Neurology/Postępy Psychiatrii i Neurologii. 2024;33(4):230-240.
 56. Sokół-Szawłowska M, Sochacka M. *MoodMon – pierwsza polska aplikacja używająca AI w chorobie afektywnej dwubiegunowej i zaburzeniach depresyjnych nawracających. Wstępna ocena z perspektywy pacjentów.* Psychiatr. Pol. 2025: 1-12. <https://doi.org/10.12740/PP/OnlineFirst/207280>

Autor korespondencyjny: Marlena Sokół-Szawłowska
e-mail: marlenasokolsz@gmail.com