ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Machine Learning-Based Quantification of (−)-*trans*-Δ-Tetrahydrocannabinol from Human Saliva Samples on a Smartphone-Based Paper Microfluidic Platform

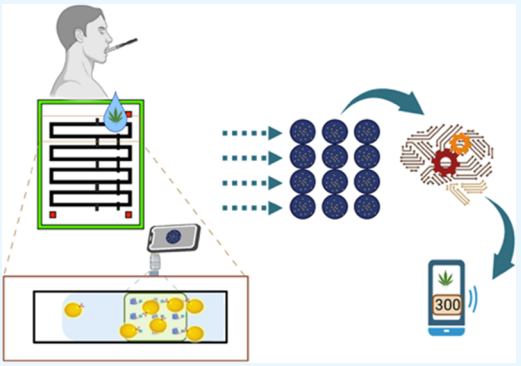
Yan Liang, Avory Zhou, and Jeong-Yeol Yoon*

Cite This: *ACS Omega* 2022, 7, 30064−30073

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: (−)-*trans*-Δ-Tetrahydrocannabinol (THC) is a major psychoactive component in cannabis. Despite the recent trends of THC legalization for medical or recreational use in some areas, many THC-driven impairments have been verified. Therefore, convenient, sensitive, quantitative detection of THC is highly needed to improve its regulation and legalization. We demonstrated a biosensor platform to detect and quantify THC with a paper microfluidic chip and a handheld smartphone-based fluorescence microscope. Microfluidic competitive immunoassay was applied with anti-THC-conjugated fluorescent nanoparticles. The smartphone-based fluorescence microscope counted the fluorescent nanoparticles in the test zone, achieving a 1 pg/mL limit of detection from human saliva samples. Specificity experiments were conducted with cannabidiol (CBD) and various mixtures of THC and CBD. No cross-reactivity to CBD was found. Machine learning techniques were also used to quantify the THC concentrations from multiple saliva samples. Multidimensional data were collected by diluting the saliva samples with saline at four different dilutions. A training database was established to estimate the THC concentration from multiple saliva samples, eliminating the sample-to-sample variations. The classification algorithms included *k*-nearest neighbor (*k*-NN), decision tree, and support vector machine (SVM), and the SVM showed the best accuracy of 88% in estimating six different THC concentrations. Additional validation experiments were conducted using independent validation sample sets, successfully identifying positive samples at 100% accuracy and quantifying the THC concentration at 80% accuracy. The platform provided a quick, low-cost, sensitive, and quantitative point-of-care saliva test for cannabis.



INTRODUCTION

Cannabis and its byproducts, widely used as psychoactive substances, have increased in their consumption due to the legalization in recent years.[1,2] While (−)-*trans*-Δ-tetrahydrocannabinol (THC) was legalized in many countries for either medical or recreational use, debates still exist about whether THC consumption would affect coordination, memory, attention, and other abilities.[3,4] Previous studies have shown that the consumption of THC would impair driving-related skills heavily and cause a significant increase in the risk of fatal accidents, especially among younger populations.[5] Despite the individual differences in the THC impact, many jurisdictions in Europe and North America highly recommended regulating THC consumption among roadside drivers and employees.[5,6]

Existing regulatory limits for THC in whole blood or plasma are varied from 1 to 5 ng/mL, but testing THC in blood is complicated, time-consuming, and environmentally restricting.[7] Most importantly, the THC concentration in blood would dramatically drop during the collection, delivery, and long screening time.[7] Previous research has demonstrated the concentration correlation between blood samples and oral fluid (saliva samples).[8,9] This correlation makes the saliva detection of THC a good alternative since it is noninvasive to collect samples and easy to handle.[10,11]

Laboratory-based THC detections include high-performance liquid chromatography (HPLC),[12] liquid chromatography-tandem mass spectrometry (LC-MS/MS),[13] gas chromatography-mass spectrometry (GC-MS),[14] and other chromatographic separation techniques.[15] These methods are sensitive and reliable, and the testing can be modified based on the regulation requirements, making these detections generally the gold standard.[5] However, complex and expensive machines are needed in addition to trained personnel. Alternatively, the affinity-based assay has popularly been used as commercial kits, e.g., competitive lateral flow immunoassays (LFIAs). However, their limit of detection (LOD) is too high (over 25 ng/mL) to meet the regulation standard of 2−4.9 ng/mL, which is

Received: May 18, 2022
Accepted: August 2, 2022
Published: August 15, 2022

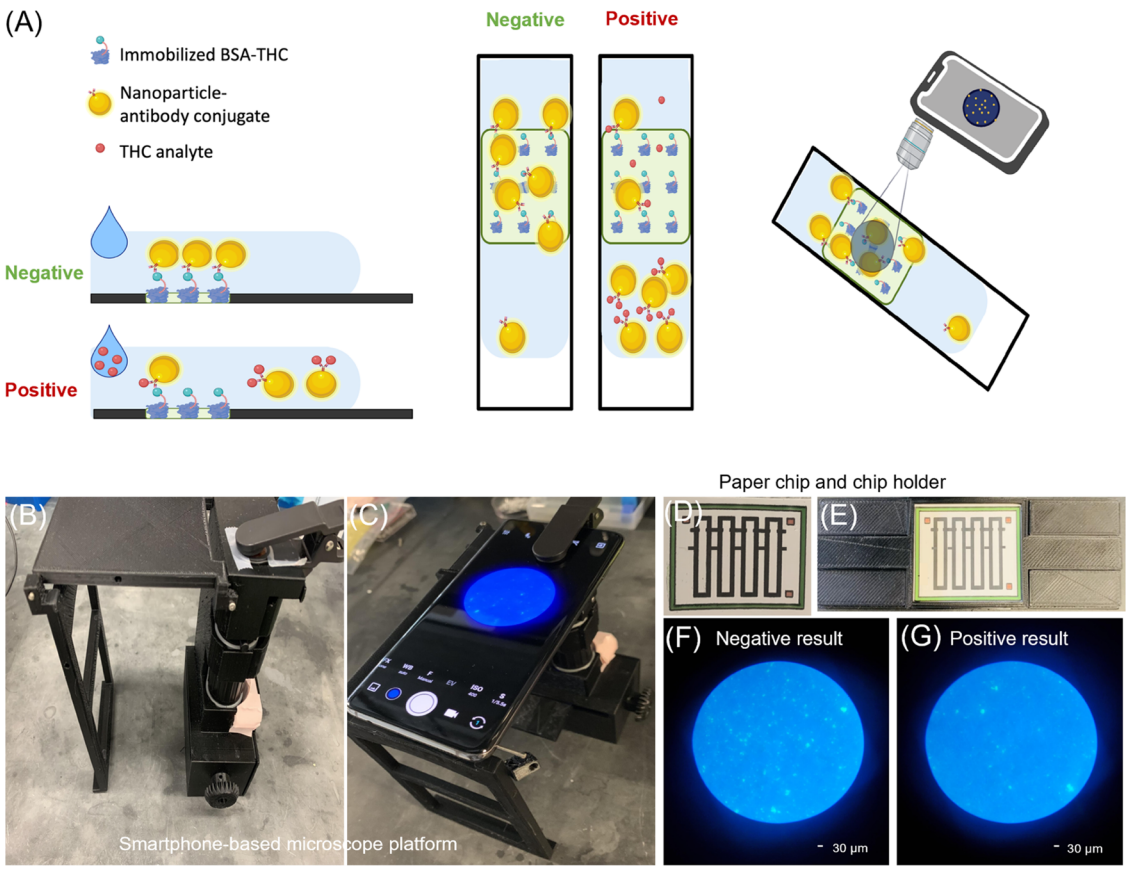


Figure 1. Detection principle and assay platform. (A) Schematics of the microfluidic competitive immunoassay for THC detection. Side and top views were shown for negative and positive tests. Antibody-conjugated nanoparticles were captured within the test zone in the negative test, while free THC analytes in the positive samples had priority to bind with antibodies, causing the reduced nanoparticle numbers remaining in the test zone. Created with BioRender.com. (B–G) Photographs of the smartphone-based microscopic platform, the paper chip, and the paper chip holder. (F–G) Smartphone microscopic images from the test zone, where panel (G) is the negative result image with a high number of captured nanoparticles and panel (F) is the positive result with low nanoparticle numbers captured in the test zone. Images were further processed with ImageJ for counting nanoparticles.

recommended by the U.S. Department of Health and Human Service.[5,16] Most importantly, current LFA detections failed to provide the THC concentration values in the samples, limiting their use in quantitative applications. Recent studies and available commercial kits are summarized in Table S1.[5,9,16−18]

For practical applications, the amount of THC in saliva or blood inevitably varies from person to person.[16] By uncovering the underlying patterns from training data, machine learning (ML) algorithms can overcome interferences generated by proteins and other molecules in saliva and blood as well as sample-to-sample variances.[19,20] Several well-known ML algorithms, such as support vector machines (SVMs), *k*-nearest neighbors (*k*-NNs), random forests (RFs), Bayesian networks, and Gaussian networks, have been used in many genomics, proteomics, and other biological applications.[21] For example, Nakano et al. trained a deep learning model and an SVM algorithm to classify oral malodor and healthy breath from oral microbiota in saliva from 16S rRNA sequences.[22] Kim et al. trained four ML models (neural network, SVM, RF, and regularized logistic regression) to differentiate healthy vs periodontitis patients, with the genomic DNA data isolated from saliva samples.[23] The average accuracies were 93% for identifying moderate-to-severe periodontitis and 78% for identifying slight periodontitis.

This work demonstrates a new biosensor platform using a microfluidic competitive immunoassay on a paper microfluidic chip and a smartphone-based fluorescence microscope. Fluorescence microscopic images were taken and further processed to count the nanoparticles captured in the test zone. Counting the number of nanoparticles in the test zone enabled the LOD to be as low as 0.5 pg/mL. Due to the variations in the amounts of proteins and other chemicals in saliva samples, the quantification of target concentration has been considered challenging. In addition, the linear range of the assay was also varied by the saliva samples, making the quantification even more difficult. In this work, we utilized ML algorithms to address these challenges. The training data set was collected by serially diluting the THC-spiked human saliva samples at four different dilutions, creating a multidimensional data set. *k*-NN, decision tree (DT), and SVM were applied to estimate the THC concentration from this multidimensional data set by finding the underlying patterns from multiple saliva samples. The entire data set was randomly split into training and test, and the THC concentrations were predicted for six different concentrations. Independent validation data sets were also prepared, and the accuracies were evaluated for predicting positive vs negative samples (binary test) and quantifying THC concentrations. This platform helps fill a gap in the existing THC detection technology and allows communities and

https://doi.org/10.1021/acsomega.2c03099
ACS Omega 2022, 7, 30064−30073

individuals to detect the THC content in saliva rapidly and economically.

EXPERIMENTAL SECTION

Antibody−Nanoparticle Conjugation. Five hundred nanometer yellow-green carboxylated polystyrene nanoparticles were purchased, with the peak excitation at 488 nm and the peak emission at 509 nm. Anti-THC monoclonal antibodies were covalently conjugated to these nanoparticles following a revised protocol from Bangs Laboratories,[24] utilizing 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDAC). Antibody conjugation was verified by monitoring the absorbance at 280 nm (Figure S1). The particle concentration was adjusted to 0.3 μg/μL by monitoring the absorbance at 488 nm.

Design and Preparation of Paper Microfluidic Chips. The microfluidic chip design was created with SolidWorks (Dassault Systèmes SE, Vélizy-Villacoublay, France), with the detailed dimensions shown in Figure S2. The test zone is labeled with two lines, which is closer to the inlet on the left. The paper chip was wax-printed using ColorQube 8580 (Xerox, Norwalk, CT) on a Sartorius CN95 nitrocellulose paper (capillary speed is 65−115 s/40 mm; thickness is 240−270 μm). It was cut into a reasonable size to fit the chip holder, and the wax was melted on a hot plate at 120 °C for 2 min. For the wax to appropriately fill the depth of the paper,[25,26] the chips were pressed flat on the hot plate with a metal block.

Before the assays, THC−bovine serum albumin (BSA) conjugates (antigen−hapten) were diluted with 75 μL of DI water and prepatterned onto each microfluidic channel (1 μL of 1.703 mg/mL solutions on each channel). These antigen−hapten conjugates (capture antigens) compete with the THC in the sample solutions to bind to the free-flowing anti-THC-conjugated nanoparticles. One microliter of THC−BSA conjugates was preloaded onto the test zone of every channel, which was labeled within two lines on the paper chip. Once the solution was dried, the passive immobilization was complete. After immobilizing the capture antigens (THC−BSA), 5 μL of antibody-conjugated nanoparticles was pipetted to the inlet of each microfluidic channel. The paper chip was placed in a dark environment for 10−15 min until dried to prevent photobleaching. Once dried, the paper chips were ready for assays (Figure 1A).

Smartphone-Based Fluorescence Microscope. In addition to the microfluidic chips, the platform set was also designed with SolidWorks and printed using a three-dimensional (3D) printer (Crealty 3D, Shenzhen, China) using poly(lactic acid) (PLA) filaments. The platform's design includes a foldable smartphone stand to provide handheld portability, a translational stage providing precise and smooth control of the movement of the paper microfluidic chip, and an opening for the power switch of the light excitation system (Figure 1B,C). In addition to the power switch, the design also allows room for installing a rechargeable battery, a two-way switch for blue excitation and a bright field lighting, and an acrylic filter card (Edmund Optics, Barrington, NJ) between the microfluidic chip and smartphone microscope. The acrylic filter card was used as a low-cost alternative to an optical filter to collect fluorescence emissions. The smartphone microscope attachment was made from a MicroFlip light-emitting diode (LED) and UV lighted pocket microscope, purchased from Carson Optical (Ronkonkoma, NY). It offers a magnification of 100−250×, as specified by the manufacturer. When using the smartphone to take pictures of the particles, the Samsung Galaxy S21 and the application Procam were used with the following settings: exposure = 1/30 s, ISO = 400, and white balance (WB-L) = 4000 K.

Paper Microfluidic Assay. During the assay, 5 μL of the sample solution was pipetted to the inlet of each channel (Figure 1D,E). Once again, the paper chip was placed in the dark until dry, and this process took 5−10 min. After the chip was dehydrated, it was placed on the paper chip holder previously described and inserted into the smartphone microscope device to take photographs of nanoparticles. For each sample, there were four repeats, and each repetition had three images for nanoparticle counting. Images were captured around the test zone, each time at a different location.

Standard Curve. Solutions of different concentrations of THC in DI water (0, 0.5, 1, 3, 5, 10, and 30 pg/mL) were prepared for building a standard curve. Since dilutions were made by dilution series, the actual concentrations could vary. Therefore, these values should represent four concentration ranges, e.g., 0, 10^{-1}, 10^{0}, and 10^{1} pg/mL. The smartphone-based microscope was used to take three images of each channel of a CN95 nitrocellulose paper chip. Images were further processed by ImageJ (U.S. National Institutes of Health, Bethesda, MD), Microsoft Excel 2019 (Microsoft Corporation, Redmond, WA) for nanoparticle counting, and GraphPad Prism 9.2.0 (GraphPad Software, San Diego, CA) for graphic plotting. The captured signals were the total nanoparticle numbers added together from the three images, and four repeats (n = 4) were applied in this experiment for each concentration of THC.

Selectivity and Cross-Reactivity Tests. The first solution set included different concentrations of CBD in DI water (0, 0.5, 1, 3, 5, 10, and 30 pg/mL, representing four concentration ranges, e.g., 0, 10^{-1}, 10^{0}, and 10^{1} pg/mL) for evaluating the selectivity. The second solution set fixed the mass concentration ratio of THC to CBD (1:1 and 1:10) but changed the THC concentrations (0−30 pg/mL), e.g., the CBD concentrations were 0−30 pg/mL (at a 1:1 mass concentration ratio) or 0−300 pg/mL (at a 1:10 mass concentration ratio). The third solution set used both THC and CBD spiked in DI water; the concentration of THC was fixed at 3 pg/mL (positive sample), while the mass concentration ratio of THC and CBD was set as 1:1, 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, and 1:10, i.e., the CBD concentrations were 3, 6, 9, 12, 15, 18, 21, 24, 27, and 30 pg/mL (representing the concentration ranges of 10^{0} and 10^{1} pg/mL); 3 pg/mL of THC with no CBD and 3 pg/mL of CBD with no THC were also prepared. The second and third sets were prepared to test the cross-reactivity. All experiments were conducted in an identical manner described above (n = 4; three images per sample).

THC-Containing Saliva Samples. NaCl solution (0.9% w/v) (standard saline) was prepared for diluting the saliva samples. Saliva samples were first diluted at 10-, 100-, 1000-, and 10,000-fold, corresponding to 10, 1, 0.1, and 0.01% saliva. A set of THC solutions were then spiked into these four saliva dilutions. THC concentrations were fixed at 0, 0.5, 1, 3, 5, 10, and 30 pg/mL for all dilutions, representing the concentration ranges of 0, 10^{-1}, 10^{0}, and 10^{1} pg/mL. With 10% saliva, these concentrations corresponded to 0−300 pg/mL in undiluted saliva, and with 1%, 0−3000 pg/mL in undiluted saliva, etc.

Data Analysis. Using ImageJ (U.S. National Institutes of Health, Bethesda, MD), the smartphone fluorescence micro-

https://doi.org/10.1021/acsomega.2c03099
ACS Omega 2022, 7, 30064−30073

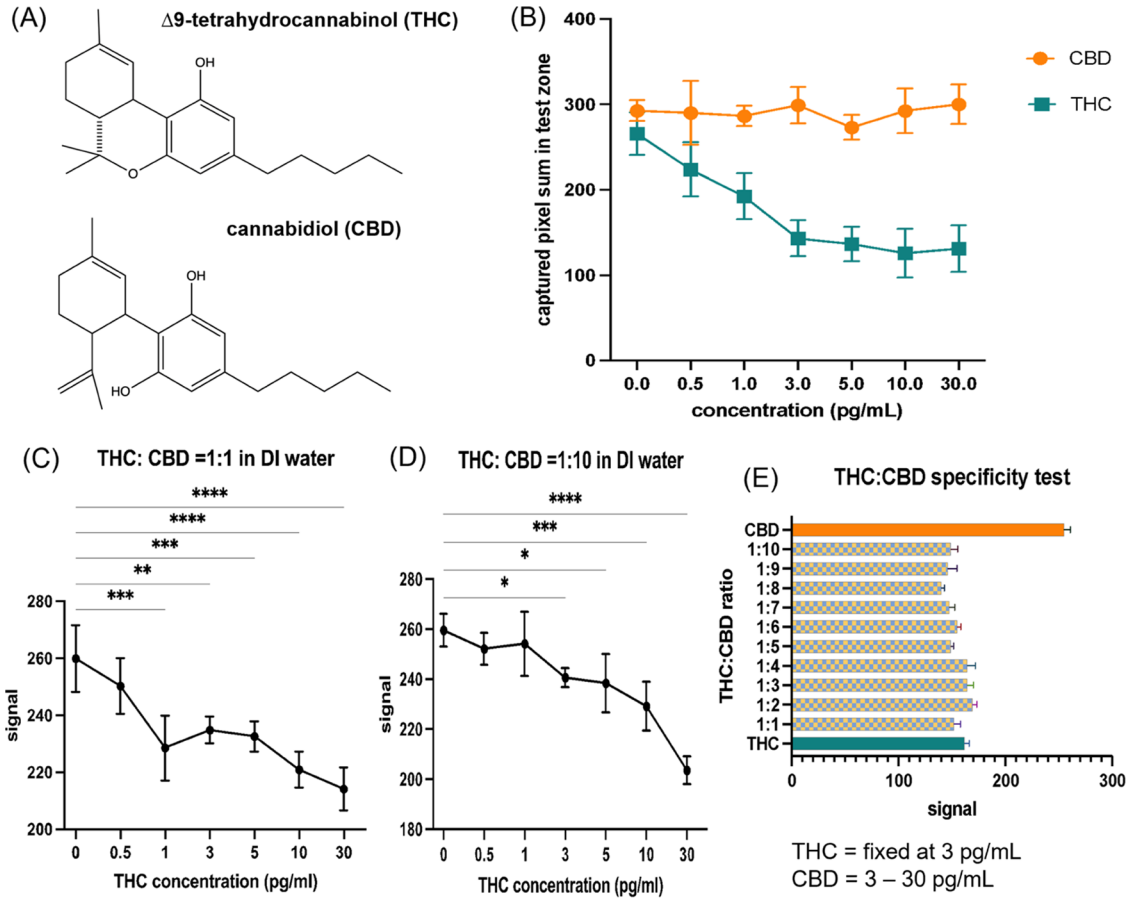


Figure 2. Specificity and cross-reactivity. (A) CBD was used for assessing specificity and cross-reactivity to THC. (B) CBD standard calibration curve in DI water did not show significant differences from the negative control, indicating the specificity. (C) The decreasing trend could still be observed over the THC concentration when THC and CBD were added at a 1:1 mass concentration ratio. LOD = 1 pg/mL. (D) The same decreasing trend could be observed at a 1:10 mass concentration ratio of THC and CBD, with the compromised LOD of 3 pg/mL. (E) The mass concentration ratio of THC and CBD was varied from 1:1 (3 pg/mL THC and 3 pg/mL CBD) to 1:10 (3 pg/mL THC and 30 pg/mL CBD), shown together with 3 pg/mL THC only and 3 pg/mL CBD only. The assay results were identically positive regardless of the CBD amount, as the THC concentration was fixed at 3 pg/mL. * means the $0.01 < p < 0.05$, which is significantly different; ** means very significant difference with $0.001 < p < 0.01$; *** means extremely significant difference with $0.0001 < p < 0.001$; **** means the difference is extremely significant with $p < 0.0001$.

scope images were processed with thresholds: 108 as the low threshold and 166 as the high threshold value (out of 255 = 8-bit). These threshold values were determined to best recapitulate the fluorescent nanoparticles compared to the benchtop fluorescence microscope images. They were auto-analyzed for the number of nanoparticles in each image using the ImageJ macro processing code (Figure 1F,G). The particle numbers were recorded, summed up from three images, and analyzed using Microsoft Excel 2019. For visual identification, the Excel sheets were further added to GraphPad Prism 9.2.0 (GraphPad Software, San Diego, CA).

For group comparison between different samples, one-way and two-way analysis of variance (ANOVA) tests were performed using GraphPad Prism 9.2.0. Through ANOVA tests, mean values between various concentrations and the control were compared to determine any significant differences. In addition, Tukey's honestly significant difference test (Tukey's HSD test) was applied.

All saliva THC detection data were later fed into an ML code, utilizing Python's Scikit-Learn library, pandas, numpy, and matplotlib.pyplot for confusion matrix analyses and visualization. *k*-NN, decision tree, and SVM were used to classify and predict the THC concentration in the multiple saliva samples. The data set was randomly split into the training set and the testing set at a ratio of 7:3. The training set was used to build the supervised learning model, while the testing set was used for accuracy evaluation. The optimization algorithm used a grid search to optimize the best parameter value, and the *k*-fold cross-validation was applied to find the best model performance. The best-performing model was utilized to obtain the confusion matrices.

RESULTS AND DISCUSSION

Detection Principle and the Assay Platform. The biosensor comprised a paper microfluidic chip and a smartphone fluorescence microscope.[27−29] Microfluidic competitive immunoassay was conducted on a paper chip to detect THC from saliva, and the nanoparticle numbers in the test zone were counted with a smartphone microscope. THC−BSA antigens were preimmobilized at the test zone as the competitive antigens, followed by adding the anti-THC-conjugated fluorescence nanoparticles to finish the chip preparation. For the detection, the sample was dropped on the paper chip channel's inlet and moved along the channel by

https://doi.org/10.1021/acsomega.2c03099
ACS Omega 2022, 7, 30064−30073

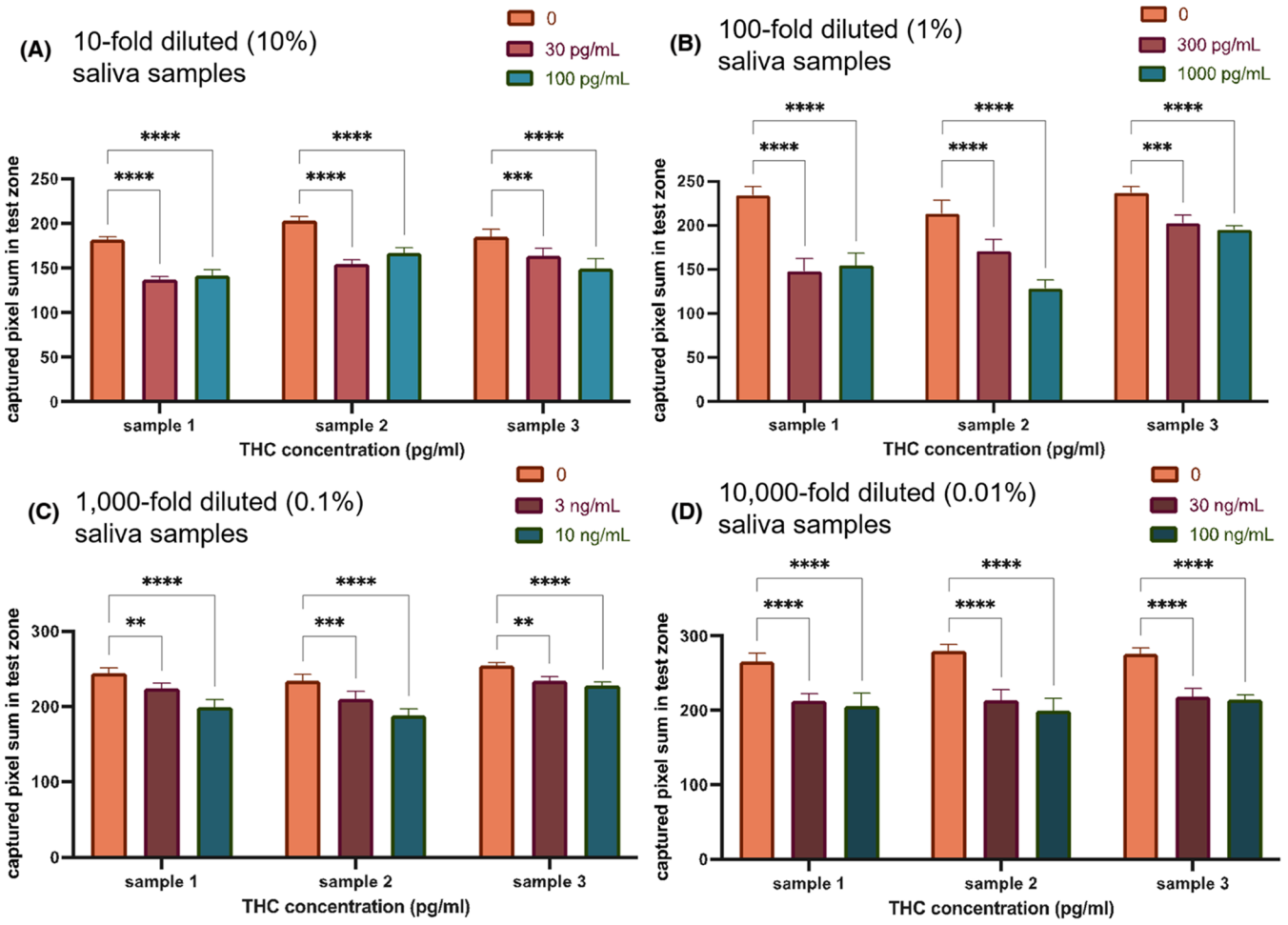


Figure 3. Qualitative test with negative samples and positive samples (with varying concentrations). Different dilution factors (10, 100, 1000, and 10,000) were applied with 0.9% w/v NaCl solution. Significant differences were shown between the negative and positive samples, verifying the feasibility of this assay in salivary detection. (A) The assay results with 10-fold dilution. The THC concentrations in the positive samples are equivalent to 30 and 100 pg/mL. (B) The same with 100-fold dilution, equivalent to 300 and 1000 pg/mL. (C) The same with 1000-fold dilution, equivalent to 3 and 10 ng/mL. (D) The same with 10,000-fold dilution, equivalent to 30 and 100 ng/mL. ** means very significant difference with $0.001 < p < 0.01$; *** means extremely significant difference with $0.0001 < p < 0.001$; **** means the difference is extremely significant with $p < 0.0001$.

capillary action. The smartphone-based fluorescence microscope with an LED light (wavelength at 460 nm) excited the fluorescence nanoparticles to generate signals (Figure 1A). With the presence of free-moving THC analytes, the antibodies on the nanoparticles were occupied and would not interact with the THC−BSA antigens immobilized in the test zone, which led to a decreasing number of fluorescence signals shown on the smartphone screen (Figure 1A,G). In the negative sample, the antibody-conjugated particles were captured at the test zone, presenting increased fluorescence signals (Figure 1A,F).

The standard curve was generated with varying THC concentrations in DI water. It shows the linearly decreasing signals in the range of 0−30 pg/mL (0−10^1 pg/mL) of THC, and the LOD for the standard curve is 0.5 pg/mL in Figure 2B, thereby establishing a proof of concept for the THC detection platform.

While Samsung Galaxy S21 was used, other smartphone brands like Samsung Galaxy S10 Lite, Apple iPhone 12, and Apple iPhone 13 were also tested. While slight color variances were observed, no significant differences were found in the pixel sums of captured nanoparticles, as shown in Figure S3.

Stability Assessments. The successful capture of antibody-conjugated nanoparticles in the test area was confirmed by taking the smartphone-based fluorescence microscopic images before and after loading the antibody−nanoparticles, where no positive controls were used (Figure S4). In addition, the size distributions of bare vs antibody-conjugated nanoparticles on paper microfluidic chips were evaluated by imaging them using a smartphone-based fluorescence microscope. The results are shown in Figure S5. In all four cases, the diameters were not substantially different from each other. This result indicates that the particles do not self-aggregate by antibody conjugation, demonstrating the stability of antibody-conjugated particles. It also shows that the particles do not aggregate by the positive sample presence.

The durability of the paper chips was also assessed over time (Figure S6). The paper chips were first loaded with THC−BSA, followed by antibody-conjugated fluorescent nanoparticles, and dried. These preloaded chips were stored at either room temperature or 4 °C for up to 10 days. THC solutions at 0, 3, and 10 ng/mL were added after days 1, 2, etc., up to day 10. Paper chips were used only once and discarded after each use. The captured pixel sums were evaluated using a

https://doi.org/10.1021/acsomega.2c03099
ACS Omega 2022, 7, 30064−30073

smartphone-based fluorescence microscope. No significant differences were observed for the prepared paper chips stored at 4 °C for both negative (0 ng/mL) and positive (3 and 10 ng/mL) samples. While a noticeable decrease could be found with the negative samples starting from day 5, no such decreases could be found with the positive samples, and successful differentiation could still be achieved between negative and positive samples.

Selectivity/Cross-Reactivity Tests with Cannabidiol (CBD). CBD shares a similar chemical structure as THC (Figure 2A) and is often present with THC together in the cannabis plants or the consumed drugs.[30] Acting as an antagonist of cannabinoid type 1 (CB1) and cannabinoid type 2 (CB2) receptors, CBD and its byproducts have been demonstrated to lack the psychoactive effects of THC and are considered relatively safe as medicines and cosmetics for neuroprotective and anti-inflammatory effects.[31−33] Thus, CBD serves as a great candidate to verify the selectivity and cross-reactivity of our THC detection method. A series of THC and CBD concentrations were spiked in DI water separately. Results in Figure 2B showed insignificant differences from blank with CBD-only solutions, while a clear declining trend was observed with THC-only solutions. This result implied the different binding kinetics between CBD and THC molecules;[34] thus, CBD would not involve in the competition assays in our detection.

To further examine whether the existence of CBD would affect the detection behavior of THC, we prepared two solution sets with various THC concentrations (0−30 pg/mL), added with 1× and 10× (mass ratios) CBD's, e.g., (0−30 and 0−300 pg/mL, respectively). Decreasing trends are shown in Figure 2C,D, indicating the low cross-activity. Additional experiments were conducted using a THC-positive sample (fixed at 3 pg/mL) mixed with different amounts of CBD (from 3 to 30 pg/mL, i.e., 1:1 to 1:10). Clear positive results were shown for all mixtures with no false-negative data (Figure 2E).

Qualitative Test from Saliva Samples. We then performed the THC detection in saliva samples instead of DI water. The saliva samples were purchased from Innovative Research and collected from different individuals. The THC-spiked saliva samples were diluted 10-fold, 100-fold, 1000-fold, and 10,000-fold to bring the THC concentrations within the detectable region. In the diluted samples, the THC concentrations varied from 0 (negative sample), 3 and 10 pg/mL (two positive samples, representing 10^0 and 10^1 pg/mL ranges). With 10-fold dilution, these positive THC concentrations corresponded to 30 and 100 pg/mL in the undiluted saliva samples. With 1000-fold dilutions, they corresponded to 3 and 10 ng/mL and 10,000-fold to 30 and 100 ng/mL. The THC-spiked saliva solutions were analyzed on the paper chip with a smartphone-based fluorescence microscope (n = 4). Dilutions were made with 0.9% saline.

As expected, significant differences between positive and negative samples can be found for all saliva dilution sets (p < 0.05 with ANOVA) (Figure 3). These results indicate our platform's successful THC qualitative detection using multiple saliva samples, with a lowered LOD to meet the regulation needs. Meanwhile, this test demonstrated that serial dilutions could achieve a wide range of target concentrations. For example, 3−10 pg/mL THC with a 10-fold dilution (Figure 3A) is equivalent to 30−100 pg/mL THC in undiluted saliva. The same THC concentrations with 10,000-fold dilution (Figure 3D) are equal to 30−100 ng/mL in undiluted saliva. The U.S. Department of Health and Human Services regulates the THC concentrations at 2−4.9 ng/mL.[16,34] Therefore, 0.1% dilution (1000-fold dilution) can be used to cover the 0.5−30 ng/mL range. In some THC zero-tolerated areas, 10% dilution (10-fold dilution) can be applied that covers the 5−300 pg/mL range. Multiple serial dilutions offer a flexible and broad range of detection. The data set can also be used to estimate the target concentration from varied human saliva samples, which is addressed later.

Quantification of THC Concentration Using Machine Learning. Although significant differences between negative and positive samples were observed in all dilutions with lowered LOD (Figure 3), quantification was still challenging as particle numbers' overall magnitude varied substantially from saliva sample to sample (Figure S7A). In addition, the recovery ratios of negative control samples (negative diluted saliva solution: negative dilution buffer) varied significantly from saliva sample to sample, as shown in Figure S7B. These variations might have been caused by the interferences of immunoglobulins, proteins, enzymes, mucins, and nitrogenous products in saliva. These molecules affected or even interfered with antibody-target binding. Although serial THC concentrations in one particular sample still resulted in a decreasing trend, the variations in recoveries are hard to be associated with either turbidity, density, viscosity, and surface tension (Figure S7C−F). However, the transmittance graph did show increased values from sample 1 to sample 12, while the R^2 between the transmittance data set and the recovery data set was only 0.6337, indicating a weak relationship (Figure S8A). The same analysis was applied for density, viscosity, and surface tension data sets, with R^2 of 0.2461, 0.1004, and 0.3091, respectively (Figure S8B−D). Thus, the difficulties observed in quantifying the THC concentration could hardly be resolved by finding a single relevant parameter, and normalization using an internal reference would be challenging and potentially impractical in clinical diagnostics.

We sought to use machine learning (ML)-based classification to address this problem.[35] We tested k-nearest neighbor (k-NN), decision tree, and support vector machine (SVM), which have popularly been used to make classifications.[36] Multidimensional data set was used to build a training data set, consisting of four different dilutions (10, 100, 1000, and 10,000; i.e., 10, 1, 0.1, and 0.01%), with three different saliva samples (Figure S9). As the concentrations of spiked THC were varied from 0 to 30 pg/mL, these four dilutions would cover a wide range of THC concentrations and subsequently the varied linear ranges of the assay. A Python code (https://scikit-learn.org/stable/) was designed to select the ideal parameters for each algorithm. The parameters of k-NN and SVM were automatically optimized within this script by a nested for loop. k-Fold cross-validation was applied to check the accuracy of our model.[37] The data set was randomly split into train vs test (7:3), and six different THC concentrations were accordingly predicted (0, 100, 300, 1000, 3000, 10,000, and 30,000 pg/mL).

The low THC concentrations (e.g., 100 and 300 pg/mL; 10^2 range) may still generate positive results in the high concentration ranges. In contrast, the high concentrations (10,000 and 30,000 pg/mL; 10^4 range) can only yield positive results in the high concentration ranges. Therefore, we first predicted 0, 100, and 300 pg/mL concentrations, ignoring the 1000- and 10,000-fold dilution data spaces as zero. If the

https://doi.org/10.1021/acsomega.2c03099
ACS Omega 2022, 7, 30064−30073

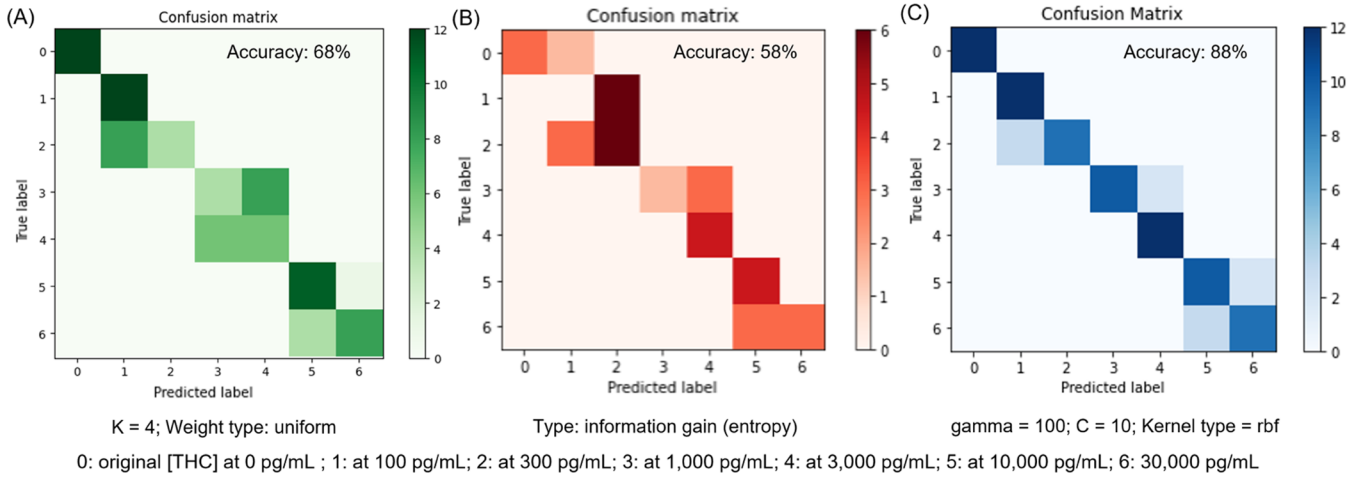
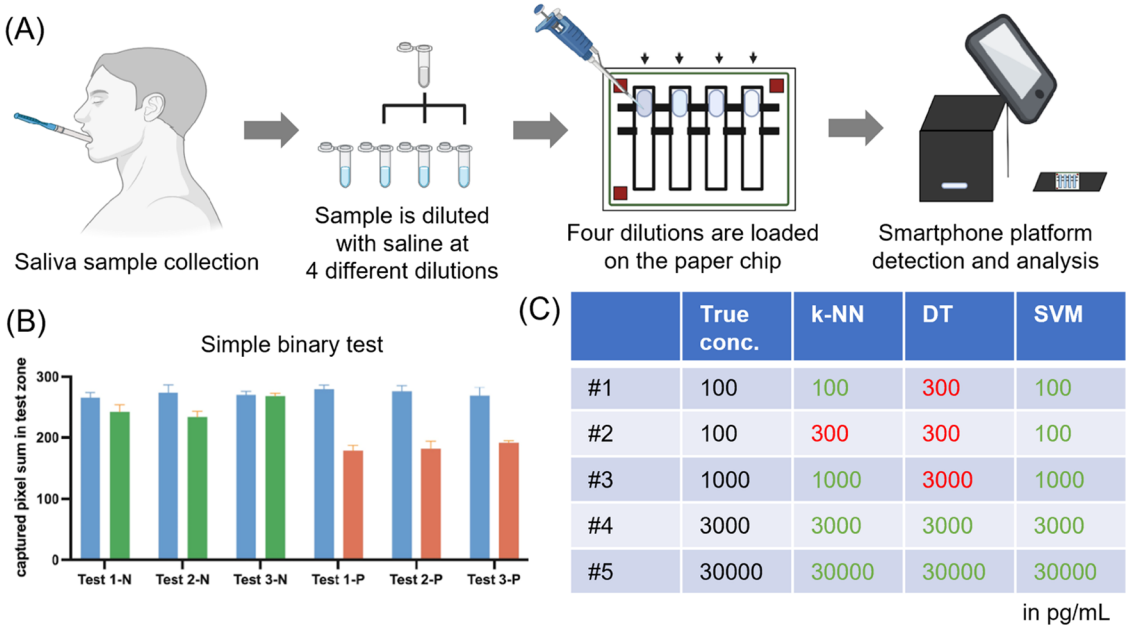


Figure 4. Machine learning-based THC quantification. Saliva samples were diluted with saline (0.9% w/v NaCl). The confusion matrix using different algorithms was shown: (A) confusion matrix with the *k*-NN model, (B) confusion matrix with the decision tree model, and (C) confusion matrix with the SVM model.



	True conc.	k-NN	DT	SVM
#1	100	100	300	100
#2	100	300	300	100
#3	1000	1000	3000	1000
#4	3000	3000	3000	3000
#5	30000	30000	30000	30000

in pg/mL

Figure 5. Validation experiments. (A) Schematics of validation experiments. Saliva samples collected from a hypothetical individual were further diluted by saline (0.9% w/v NaCl). THC was spiked into three different human saliva samples. Saline (as a negative control) and three diluted samples were added to the prepared paper chip and were left to dry. The smartphone-based fluorescence microscope captured three images from the test zone and counted the captured nanoparticles. If one of the dilutions gave the signals significantly lower than the negative control, it meant THC was present, i.e., positive. It was negative if all dilutions gave the signals not significantly different from the negative control. Created with BioRender.com. (B) Binary qualitative test with three negative and three positive samples. (C) Validation of ML models for quantification detection. DT = decision tree. Five samples were prepared, diluted into 10-, 100-, 1000-, and 10,000-fold using saline, and analyzed using the workflow shown in Figure S10. The results shown in Figure 4 were used as a training set. The ML model's predictions were compared with the true concentration values.

prediction gave the result as either 0 or 100 pg/mL, we considered it a correct prediction and stopped further analysis. If the prediction was 300 pg/mL, there was a possibility that the actual concentration could be higher. In such a case, a further prediction of 0, 1000, and 3000 pg/mL was conducted. If the prediction was 0, the actual concentration was lower than 1000 pg/mL, and the result of 300 pg/mL was accepted. It was also accepted as the actual concentration if it was 1000 pg/mL. If it was 3000 pg/mL, we proceed to the third step, predicting 0, 10,000, and 30,000 pg/mL concentrations. This prediction workflow can be found in Figure S10.

The results were shown as confusion matrices in Figure 4. The numbers 0−6 represented three different original concentrations in undiluted saliva (0, 100, 300, 1000, 3000, 10,000, and 30,000 pg/mL); the predicted label was shown on the horizontal axis, and a true label was on the vertical axis. The *k*-NN model has an accuracy of 68% for training and testing. As a lazy learning model (instance-based learning), *k*-NN does not go through and learn from the training data set until making real-time predictions, which allows adding new data seamlessly without impacting the algorithm's accuracy.[38] The accuracy for the decision tree was low (58%), since the

https://doi.org/10.1021/acsomega.2c03099
ACS Omega 2022, 7, 30064−30073

decision tree is usually greedy and deterministic, forcing the consideration of all possible outcomes of a decision and tracing each path to a conclusion, which tends to overfit.[38] SVM achieved the highest accuracy of 88%, showing the ability to increase class separation and minimize certain prediction errors.[39] For the 12% mispredicted data, they were all predicted to the neighboring concentrations, e.g., 300−100 (both in 10^2 range), 1000−3000 (both in 10^3 range), and 10,000−30,000 (both in 10^4 range) pg/mL, i.e., within an order of magnitude. As discussed in the Experimental Section, THC solutions were prepared by dilution series to reach pg/mL range. (All assays were conducted with the THC concentrations from 0 to 30 pg/mL with varying saliva concentrations; e.g., 10−30 pg/mL in 1% saliva corresponded to 1000−3000 pg/mL in undiluted saliva.) Therefore, the actual THC concentrations could become uncertain in this pg/mL range, and it should represent the concentration ranges of 10^0, 10^1, 10^2 pg/mL, etc., rather than the exact concentrations. This argument could explain the misprediction of the neighboring concentration.

Validation Experiments: Binary Qualitative Test and Quantification with ML Models. To further demonstrate the practicality of our THC detection, we created a scenario of identifying positive vs negative results from unknown samples. (1) A saliva sample is collected from an individual. (2) Saline (0.9% w/v NaCl) is added to create multiple dilutions (10-, 100-, 1000-, and 10,000-fold, e.g., 10, 1, 0.1, and 0.01%). A pure saline sample is additionally tested as a negative control. (3) Solutions are added separately to the prepared (THC−BSA-immobilized) paper chip channels. It takes several minutes for the sample to dry completely. (4) The paper chip is placed on the chip holder and inserted into the smartphone fluorescence microscope platform. The smartphone captures three different images from the test zone (within two labeled lines on the paper chip). ImageJ isolates the fluorescent nanoparticles and counts them (Figure 5A). (5) Binary qualitative test is conducted, i.e., differentiating positive from negative samples. The signal from a saline solution should be significantly higher than those from positive samples. The binary assay results are shown in Figure 5B, using these independent validation data sets. The accuracy was 100%. (6) Quantification is conducted using the ML models. The entire database described in the previous section is used as a training set. Five samples were created: 100, 300, 1000, 3000, and 30,000 pg/mL THC spiked to saliva samples and further diluted into 10-, 100-, 1000-, and 10,000-fold using the saline buffer. The equivalent concentrations in the undiluted saliva covered the typical THC concentrations in human blood and saliva. Since our method was more sensitive in the low concentration range, we would start from 0, 100, and 300 pg/mL—the workflow described previously (shown in Figure 4C) was also used for this independent validation. As shown in Figure 5C, the *k*-NN model successfully predicted the concentrations as 100, 100, 1000, 3000, and 30,000 pg/mL (or 0.1, 0.1, 1, 3, and 10 ng/mL) in the undiluted saliva. The accuracy was 80% since sample #2 with 300 pg/mL concentration was wrongly predicted as 100 pg/mL. Again, it was still within the same order of magnitude (10^2 pg/mL) and could be explained by the uncertainty of the dilution series. The accuracy for the decision tree prediction was only 40%, with the correct predictions at 3000 and 30,000 pg/mL (10^3 and 10^4 pg/mL ranges). The SVM model worked the best with the current data set, with an accuracy of 100% (Figure 5C). Further research may be necessary with additional clinical samples to reinforce the training database.

CONCLUSIONS

This study reported a microfluidic competitive immunoassay for THC detection on a paper microfluidic chip. A smartphone-based fluorescence microscope provided a convenient way to count the fluorescent nanoparticle numbers in the test zone, which significantly improved THC detection's analytical sensitivity (LOD). The assay time was 10 min. Meanwhile, the high selectivity and negligible cross-reactivity were verified using CBD, another significant component in cannabis, to minimize the false results. 0.9% w/v NaCl saline solution was chosen as the dilution buffer. Significant differences from all dilution sets demonstrated the validation in salivary detection and provided an adjustable detection range to meet different needs. The ML model quantified the THC concentration from multiple saliva samples despite the individual variances and interferences from proteins and other salivary substances. The SVM algorithm model accurately predicted six different THC concentrations—88% with the train−test split test and 80% with the independent validation data set. The accuracy in predicting positive vs. negative samples was 100% with the independent validation data set. This THC detection platform can serve as a promising tool for field applications, meeting qualitative and quantitative requirements.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c03099.

Verification of antibody−nanoparticle conjugation (Figure S1); channel design and dimensions (Figure S2); comparisons of raw and processed images over different brands of smartphones (Figure S3); smartphone-based fluorescence microscopic images of paper microfluidic chips (preloaded with THC−BSA) before and after loading antibody-conjugated nanoparticles (Figure S4); size distributions of bare vs antibody-conjugated nanoparticles on paper microfluidic chips, as imaged by a smartphone-based fluorescence microscope (Figure S5); durability test of the paper chips, preloaded with THC−BSA and antibody−nanoparticles and stored at room temperature and 4 °C (Figure S6); characteristics of saline- and DI water diluted human saliva samples (Figure S7); correlation analysis between the saliva samples' physical properties and recovery: (A) transmittance, (B) density, (C) viscosity, and (D) surface tension (Figure S8); ML prediction of THC concentration (Figure S9); workflow of ML-based THC quantification (Figure S10) (PDF)

AUTHOR INFORMATION

Corresponding Author

Jeong-Yeol Yoon − *Department of Chemistry and Biochemistry, The University of Arizona, Tucson, Arizona 85721, United States; Department of Biomedical Engineering, The University of Arizona, Tucson, Arizona 85721, United States;* orcid.org/0000-0002-9720-6472; Email: jyyoon@arizona.edu

https://doi.org/10.1021/acsomega.2c03099
ACS Omega 2022, 7, 30064−30073

Authors

Yan Liang − *Department of Chemistry and Biochemistry, The University of Arizona, Tucson, Arizona 85721, United States*

Avory Zhou − *Department of Biomedical Engineering, The University of Arizona, Tucson, Arizona 85721, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c03099

Author Contributions

Y.L.: conceptualization, data curation, formal analysis, investigation, methodology, resources, software, validation, visualization, and writing—original draft; A.Z.: data curation, formal analysis, investigation, methodology, resources, validation, and writing—review and editing; and J.-Y.Y.: conceptualization, formal analysis, funding acquisition, investigation, project administration, supervision, validation, visualization, and writing—review and editing.

Funding

This work was supported by the University of Arizona Test All Test Smart program.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge Alexander S. Day and Katelyn Sosnowski at the University of Arizona for their help in the data analysis discussion. The authors also acknowledge Lane E. Breshears, Brandon T. Nguyen, Patarajarin Akarapipad, and Sangsik Kim, all at the University of Arizona, for their help in designing and fabricating the smartphone-based fluorescence microscope.

■ ABBREVIATIONS

CCR2, CC chemokine receptor 2; CCL2, CC chemokine ligand 2; CCR5, CC chemokine receptor 5; TLC, thin-layer chromatography; THC, (−)-*trans*-Δ-tetrahydrocannabinol; CBD, cannabidiol; ML, machine learning; *k*-NN, *k*-nearest neighbor; SVM, support vector machine

■ REFERENCES

(1) Hazekamp, A.; Fischedick, J. T. Cannabis - from cultivar to chemovar. *Drug Test. Anal.* **2012**, *4*, 660−667.

(2) Smart, R.; Caulkins, J. P.; Kilmer, B.; Davenport, S.; Midgette, G. Variation in cannabis potency and prices in a newly legal market: evidence from 30 million cannabis sales in Washington state. *Addiction* **2017**, *112*, 2167−2177.

(3) Prashad, S.; Filbey, F. M. Cognitive motor deficits in cannabis users. *Curr. Opin. Behav. Sci.* **2017**, *13*, 1−7.

(4) Ramaekers, J. G.; Mason, N. L.; Kloft, L.; Theunissen, E. L. The why behind the high: determinants of neurocognition during acute cannabis exposure. *Nat. Rev. Neurosci.* **2021**, *22*, 439−454.

(5) Thapa, D.; Samadi, N.; Patel, N.; Tabatabaei, N. Thermographic detection and quantification of THC in oral fluid at unprecedented low concentrations. *Biomed. Opt. Express* **2020**, *11*, 2178−2190.

(6) Oiye, É. N.; Ribeiro, M. F. M.; Ferreira, B.; Botelho, R. C. B.; de Oliveira, M. F. Disposable 3D Printed electrode for the electrochemical detection of delta-9-tetrahydrocannabinol in aqueous solution and 11-nor-9-carboxy-tetrahydrocannabinol in saliva. *Braz. J. Forensic Sci., Med. Law Bioethics* **2020**, *9*, 521−533.

(7) Hartman, R. L.; Brown, T. L.; Milavetz, G.; Spurgin, A.; Gorelick, D. A.; Gaffney, G. R.; Huestis, M. A. Effect of blood collection time on measured δ9-tetrahydrocannabinol concentrations: implications for driving interpretation and drug policy. *Clin. Chem.* **2016**, *62*, 367−377.

(8) Huestis, M. A.; Cone, E. J. Relationship of Δ9-tetrahydrocannabinol concentrations in oral fluid and plasma after controlled administration of smoked cannabis. *J. Anal. Toxicol.* **2004**, *28*, 394−399.

(9) Wanklyn, C.; Burton, D.; Enston, E.; Bartlett, C.-A.; Taylor, S.; Raniczkowska, A.; Black, M.; Murphy, L. Disposable screen printed sensor for the electrochemical detection of delta-9-tetrahydrocannabinol in undiluted saliva. *Chem. Cent. J.* **2016**, *10*, No. 1.

(10) Lee, D.; Huestis, M. A. Current knowledge on cannabinoids in oral fluid. *Drug Test. Anal.* **2014**, *6*, 88−111.

(11) Samadi, N.; Thapa, D.; Salimi, M.; Parkhimchyk, A.; Tabatabaei, N. Low-cost active thermography using cellphone infrared cameras: from early detection of dental caries to quantification of THC in oral fluid. *Sci. Rep.* **2020**, *10*, No. 7857.

(12) Jatoi, A.; Windschitl, H. E.; Loprinzi, C. L.; Sloan, J. A.; Dakhil, S. R.; Mailliard, J. A.; Pundaleeka, S.; Kardinal, C. G.; Fitch, T. R.; Krook, J. E.; Novotny, P. J.; Christensen, B. Dronabinol versus megestrol acetate versus combination therapy for cancer-associated anorexia: a north central cancer treatment group study. *J. Clin. Oncol.* **2002**, *20*, 567−573.

(13) Aizpurua-Olaizola, O.; Omar, J.; Navarro, P.; Olivares, M.; Etxebarria, N.; Usobiaga, A. Identification and quantification of cannabinoids in *Cannabis sativa* L. plants by high performance liquid chromatography-mass spectrometry. *Anal. Bioanal. Chem.* **2014**, *406*, 7549−7560.

(14) Hazekamp, A.; Simons, R.; Peltenburg-Looman, A.; Sengers, M.; van Zweden, R.; Verpoorte, R. Preparative isolation of cannabinoids from *Cannabis sativa* by centrifugal partition chromatography. *J. Liq. Chromatogr. Relat. Technol.* **2004**, *27*, 2421−2439.

(15) Pourseyed Lazarjani, M.; Torres, S.; Hooker, T.; Fowlie, C.; Young, O.; Seyfoddin, A. Methods for quantification of cannabinoids: a narrative review. *J. Cannabis Res.* **2020**, *2*, No. 35.

(16) Chand, R.; Mittal, N.; Srinivasan, S.; Rajabzadeh, A. R. Upconverting nanoparticle clustering based rapid quantitative detection of tetrahydrocannabinol (THC) on lateral-flow immunoassay. *Analyst* **2021**, *146*, 574−580.

(17) Majak, D.; Fan, J.; Kang, S.; Gupta, M. Delta-9-tetrahydrocannabinol (Δ9-THC) sensing using an aerosol jet printed organic electrochemical transistor (OECT). *J. Mater. Chem. B* **2021**, *9*, 2107−2117.

(18) Stevenson, H.; Bacon, A.; Joseph, K. M.; Gwandaru, W. R. W.; Bhide, A.; Shnkhala, D.; Dhamu, V. N.; Prasad, S. A rapid response electrochemical biosensor for detecting THC in saliva. *Sci. Rep.* **2019**, *9*, No. 12701.

(19) Doğan, V.; Yüzer, E.; Kılıç, V.; Şen, M. Non-enzymatic colorimetric detection of hydrogen peroxide using a μPAD coupled with a machine learning-based smartphone app. *Analyst* **2021**, *146*, 7336−7344.

(20) Chung, S.; Breshears, L. E.; Perea, S.; Morrison, C. M.; Betancourt, W. Q.; Reynolds, K. A.; Yoon, J.-Y. Smartphone-based paper microfluidic particulometry of norovirus from environmental water samples at the single copy level. *ACS Omega* **2019**, *4*, 11180−11188.

(21) Larrañaga, P.; Calvo, B.; Santana, R.; Bielza, C.; Galdiano, J.; Inza, I.; Lozano, J. A.; Armañanzas, R.; Santafé, G.; Pérez, A.; Robles, V. Machine learning in bioinformatics. *Briefings Bioinf.* **2006**, *7*, 86−112.

(22) Nakano, Y.; Suzuki, N.; Kuwata, F. Predicting oral malodour based on the microbiota in saliva samples using a deep learning approach. *BMC Oral Health* **2018**, *18*, No. 128.

(23) Kim, E.-H.; Kim, S.; Kim, H.-J.; Jeong, H.; Lee, J.; Jang, J.; Joo, J.-Y.; Shin, Y.; Kang, J.; Park, A. K.; Lee, J.-Y.; Lee, S. Prediction of chronic periodontitis severity using machine learning models based on salivary bacterial copy number. *Front. Cell. Infect. Microbiol.* **2020**, *10*, No. 571515.

(24) Bangs Laboratories. Tech Note 205: Covalent Coupling. https://www.bangslabs.com/sites/default/files/imce/docs/TechNote%20205%20Web.pdf (accessed May 14, 2022).

(25) Qi, J.; Li, B.; Wang, X.; Zhang, Z.; Wang, Z.; Han, J.; Chen, L. Three-dimensional paper-based microfluidic chip device for multiplexed fluorescence detection of Cu^{2+} and Hg^{2+} ions based on ion imprinting technology. *Sens. Actuat. B* **2017**, *251*, 224−233.
(26) Han, J.; Qi, A.; Zhou, J.; Wang, G.; Li, B.; Chen, L. Simple way to fabricate novel paper-based valves using plastic comb binding spines. *ACS Sens.* **2018**, *3*, 1789−1794.
(27) Li, B.; Yu, L.; Qi, J.; Fu, L.; Zhang, P.; Chen, L. Controlling capillary-driven fluid transport in paper-based microfluidic devices using a movable valve. *Anal. Chem.* **2017**, *89*, 5707−5712.
(28) Zhu, G.; Yin, X.; Jin, D.; Zhang, B.; Gu, Y.; An, Y. Paper-based immunosensors: Current trends in the types and applied detection techniques. *TrAC, Trends Anal. Chem.* **2019**, *111*, 100−117.
(29) Guo, X.; Zong, L.; Jiao, Y.; Han, Y.; Zhang, X.; Xu, J.; Li, L.; Zhang, C.; Liu, Z.; Ju, Q.; Liu, J.; Xu, Z.; Yu, H.-D.; Huang, W. Signal-enhanced detection of multiplexed cardiac biomarkers by a paper-based fluorogenic immunodevice integrated with zinc oxide nanowires. *Anal. Chem.* **2019**, *91*, 9300−9307.
(30) Czégény, Z.; Nagy, G.; Babinszki, B.; Bajtel, Á.; Sebestyén, Z.; Kiss, T.; Csupor-Löffler, B.; Tóth, B.; Csupor, D. CBD, a precursor of THC in e-cigarettes. *Sci. Rep.* **2021**, *11*, No. 8951.
(31) Bara, A.; Ferland, J.-M. N.; Rompala, G.; Szutorisz, H.; Hurd, Y. L. Cannabis and synaptic reprogramming of the developing brain. *Nat. Rev. Neurosci.* **2021**, *22*, 423−438.
(32) Pertwee, R. G. The diverse CB1 and CB2 receptor pharmacology of three plant cannabinoids: delta9-tetrahydrocannabinol, cannabidiol and delta9-tetrahydrocannabivarin. *Braz. J. Pharmacol.* **2008**, *153*, 199−215.
(33) Pertwee, R. G. Pharmacological and therapeutic targets for δ9 tetrahydrocannabinol and cannabidiol. *Euphytica* **2004**, *140*, 73−82.
(34) Lee, D.; Milman, G.; Barnes, A. J.; Goodwin, R. S.; Hirvonen, J.; Huestis, M. A. Oral fluid cannabinoids in chronic, daily cannabis smokers during sustained, monitored abstinence. *Clin. Chem.* **2011**, *57*, 1127−1136.
(35) Wang, K.; Yang, J.; Xu, H.; Cao, B.; Qin, Q.; Liao, X.; Wo, Y.; Jin, Q.; Cui, D. Smartphone-imaged multilayered paper-based analytical device for colorimetric analysis of carcinoembryonic antigen. *Anal. Bioanal. Chem.* **2020**, *412*, 2517−2528.
(36) Schackart, K. E.; Yoon, J.-Y. Machine learning enhances the performance of bioreceptor-free biosensors. *Sensors* **2021**, *21*, No. 5519.
(37) Guo, X.; Khalid, M. A.; Domingos, I.; Michala, A. L.; Adriko, M.; Rowel, C.; Ajambo, D.; Garrett, A.; Kar, S.; Yan, X.; Reboud, J.; Tukahebwa, E. M.; Cooper, J. M. Smartphone-based DNA diagnostics for malaria detection using deep learning for local decision support and blockchain technology for security. *Nat. Electron.* **2021**, *4*, 615−624.
(38) Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: New York, 2001.
(39) Xia, Y. Correlation and Association Analyses in Microbiome Study Integrating Multiomics in Health and Disease. In *The Microbiome in Health and Disease*; Sun, J., Ed.; Progress in Molecular Biology and Translational Science; Elsevier Inc., 2020; Vol. *171*, pp 309−491.

https://doi.org/10.1021/acsomega.2c03099
ACS Omega 2022, 7, 30064−30073