IdentiCan: The App That Detects Breast, Lung, and Skin Cancer

Cancer in Our Community

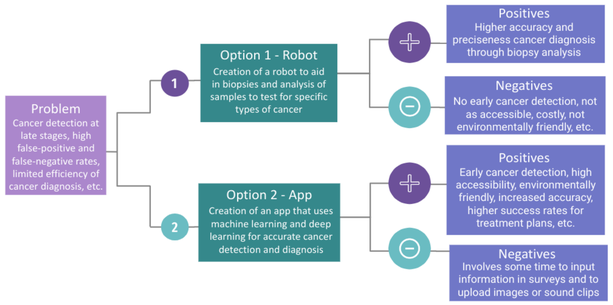
There have been many cases of cancer that we have observed in our community, including those of friends and families. After we explored the topic of cancer, we found that there were many people dying, especially from the detection of cancer at a later stage, and possibly even due to a false result when screening for cancer.

Research

After further research, we found that there were not any online tools for patients to screen themselves for basic cancer detection and diagnosis. Additionally, there were not any apps or websites for doctors to easily input patient information and scans to provide an accurate diagnosis and specialized treatment plan based on reliable algorithms.

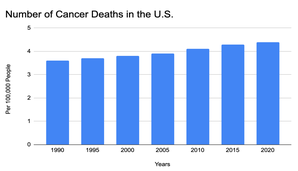
Creation of Solution

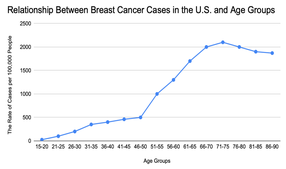
Through the integration of artificial intelligence like machine learning and deep learning, as well as a personalized survey and other features, we found that we could program our own mobile app to detect cancer at an earlier stage, increase accuracy, provide more specialized treatment plans for each patient, etc.



Significance of Our Research

An observed increase in cancer cases around the world has led to the need for a digitized tool for cancer diagnosis.
IdentiCan minimizes the mortality rate of cancer through its detection at an earlier stage as well as through increased success rates of treatment plans, which are generated by artificial intelligence algorithms, to ensure a higher success of cancer treatment.
IdentiCan allows for the prediction of the future growth of tumors in terms of rate of growth and possible spreading, based on the aggressiveness of the cancer.
IdentiCan provides higher accuracy rates, more efficient diagnosis methods (reduction in cost required and resources), decreased false-positive and negative rates, etc., through algorithms & machine learning.





	Cigarette smoking:	Secondhand smoking:	Radiation exposure:	Occupational exposure to lung carcinogens:	Air pollution:
Magnitude of effect:	Increased risk; very large	Increased risk; small magnitude	Increased risk; depending on the amount of exposure	Increased risk; large magnitude	Increased risk; varies depending on the exposure
Internal validity:	Good	Good	Fair	Good	Good
Consistency:	Good	Good	Good	Good	Good
External Validity:	Good	Good	Good	Good	Good

Engineering Goals

- Log and conduct at least 6,000 trials per each type of cancer to develop a high accuracy
- Incorporate at least 5 different global languages to maximize user accessibility
- Incorporate more advanced fractal geometry
- Increase the number of features available in IdentiCan
- An overall accuracy of at least 90 percent in cancer detection and diagnosis for IdentiCan
- Improve the accuracy of sound classification by inputting at least 500 more sound clips
- Develop a prototype of IdentiCan on Android Studio

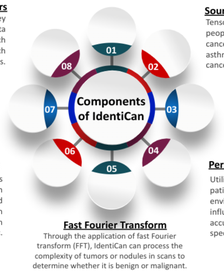


Image Classification
Machine learning model trained using 24,000 images were used to train the model, along with 2,000 separate images used to cross validate for tumors, nodules, moles, etc.

Doctor and Patient Users
Projected users are doctors and patients. Key difference: doctor side includes the collection of data for various patients and the ability store data for each patient to review personal information and track each patient's diagnosis.

Messaging Platform
Allows doctors and patients are able to communicate regarding the progression of cancer treatment, reminders of upcoming appointments, emergencies, etc.

Personalized Results and Treatment Plans
IdentiCan provides a specific diagnosis and connects the users (on the patient side) to doctors near them for further consultation. Based on the predicted future rate of growth and other factors, IdentiCan also provides personalized treatment plans, with information regarding cost, basic overview, etc.

Sound Classification
Tensorflow was trained with sound clips of people with varying conditions to identify lung cancer to differentiate between someone with asthma, normal breathing, or possibly lung cancer.

Fractal Geometry
Applied to predict the rate of growth for cancer, and to allow for the most personalized treatment plan for each patient's cancer.

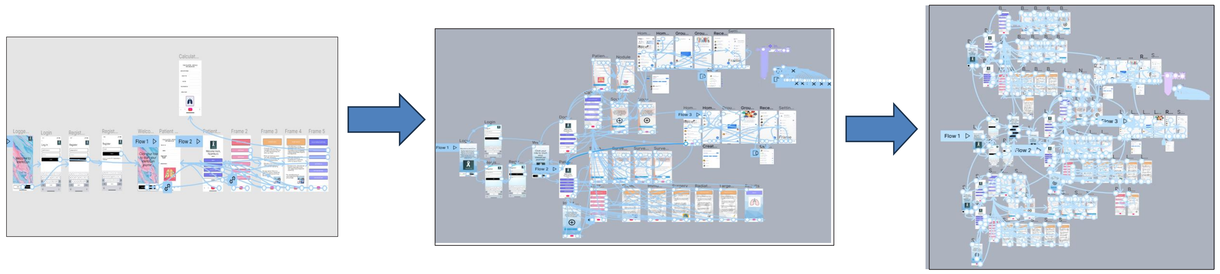
Personalized Surveys
Utilizes the input of user information about the patient's medical history, family history, environmental factors, and other factors that influence the likeliness of cancer. To increase the accuracy, machine learning allow for more specific and accurate cancer diagnosis.

Fast Fourier Transform
Through the application of fast Fourier transform (FFT), IdentiCan can process the complexity of tumors or nodules in scans to determine whether it is benign or malignant.

Methods

- We researched the different types of symptoms experienced, risk factors, treatments, previous studies and results such as HPPI-TOFMS (high-pressure photon ionization time-of-flight mass spectrometry).
- We have run various trials, which all account for patients of various backgrounds, different medical histories, and complications that may arise.
- We planned the app digitally and on paper to gain an understanding of what exactly we wanted.
- We have built a simple digital prototype in Figma and a second prototype was developed on Xcode, where we have used coding languages such as JavaScript, Kotlin, and Python.
- The third version of our app is in Android Studio, and this is the prototype that we are looking to release in the future.

Design Evolution: Prototypes 1.1 - 1.3



Procedures

Machine Learning Procedure Example:

1. Train the machine using approximately a minimum of 500 pictures in each class (different types of each cancer) and approximately 8,000 pictures each for breast, skin, and lung cancer
2. Experiment with numbers of iterations and augmentations to ensure highest accuracy in both image and sound classifications and record data for each trial
3. Cross validate machine learning model using images with varying types, stages, and progression of cancer
4. Run simulations combining the machine learning process with the survey in the view of IdentiCan users to generate examples of cancer diagnosis, treatment plans, etc.
5. Repeat using rare cancer cases, patients who are minorities, different types of cancer scans, etc.

Treatment Plan Success Rate Procedure Example:

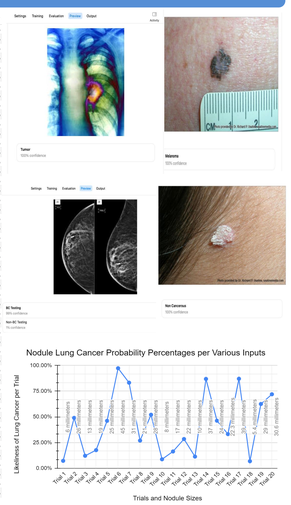
1. Analyze inputted data from the survey, images, sound clips (if applicable), generated diagnosis, etc.
2. Apply fractal geometry and FFT to predict growth and rate of growth of tumor, nodule, mole, etc., in the future
3. Identify best treatment plan based on predicted growth and aggression, diagnosis, and other factors
4. Run series of simulations to identify the success rate of various treatment plans generated by IdentiCan on different patients examples

Results

Date	Trial Number	Number of Images Used to Train Machine	Benign or Malignant Outcome	Number of Iterations	Accuracy
September 4, 2022	1	5	Unclassified	0	0%
September 5, 2022	2	25	Squamous Cell Carcinoma	2	1.32%
September 6, 2022	3	90	Unclassified	6	2.71%
September 7, 2022	4	145	Benign	9	5.52%
September 8, 2022	5	285	Benign	14	5.90%
September 10, 2022	6	370	Malignant	24	7.84%
September 11, 2022	7	425	Melanoma	48	7.95%
September 12, 2022	8	605	Malignant	28	8.47%
September 13, 2022	9	735	Unclassified	52	10.12%
September 14, 2022	10	920	Basal Cell Carcinoma	95	11.72%
September 15, 2022	11	1085	Malignant	83	13.49%
September 16, 2022	12	1195	Benign	107	15.16%
September 17, 2022	13	1305	Unclassified	118	16.32%
September 18, 2022	14	1495	Malignant	123	21.03%
September 20, 2022	15	1560	Malignant	100	15.39%
September 21, 2022	16	1615	Squamous Cell Carcinoma	125	22.98%
September 23, 2022	17	1645	Malignant	135	25.36%
September 25, 2022	18	1785	Benign	150	27.14%
September 26, 2022	19	1800	Melanoma	152	28.62%
September 28, 2022	20	1820	Benign	173	30.06%
September 29, 2022	21	1895	Unclassified	245	32.40%
September 30, 2022	22	1915	Melanoma	310	34.66%
October 1, 2022	23	1970	Malignant	328	36.79%
October 4, 2022	24	2045	Benign	379	39.27%
October 5, 2022	25	2175	Unclassified	472	42.65%
October 6, 2022	26	2290	Unclassified	506	44.58%
October 8, 2022	27	2355	Benign	432	40.32%
October 9, 2022	28	2485	Malignant	468	42.74%
October 11, 2022	29	2530	Basal Cell Carcinoma	506	44.82%
October 12, 2022	30	2540	Malignant	510	45.05%
October 13, 2022	31	2610	Benign	546	46.59%
October 15, 2022	32	2730	Benign	572	47.34%
October 16, 2022	33	2785	Melanoma	591	48.13%
October 18, 2022	34	2845	Benign	603	49.64%
October 19, 2022	35	2960	Unclassified	618	50.37%
October 21, 2022	36	3030	Basal Cell Carcinoma	624	51.09%
October 22, 2022	37	3105	Benign	693	52%
October 24, 2022	38	3190	Malignant	647	47.14%
October 25, 2022	39	3285	Malignant	610	45.78%
October 27, 2022	40	3305	Benign	708	48.53%
October 28, 2022	41	3455	Squamous Cell Carcinoma	716	48.96%
October 30, 2022	42	3510	Malignant	732	50.14%
November 1, 2022	43	3585	Benign	748	51.04%
November 3, 2022	44	3670	Benign	730	49.26%
November 4, 2022	45	3805	Benign	779	52.66%
November 6, 2022	46	3820	Unclassified	773	51.74%
November 7, 2022	47	3895	Squamous Cell Carcinoma	802	53.92%
November 9, 2022	48	3930	Benign	809	54.86%
November 10, 2022	49	3995	Melanoma	817	55.03%
November 14, 2022	50	4010	Malignant	832	57.83%

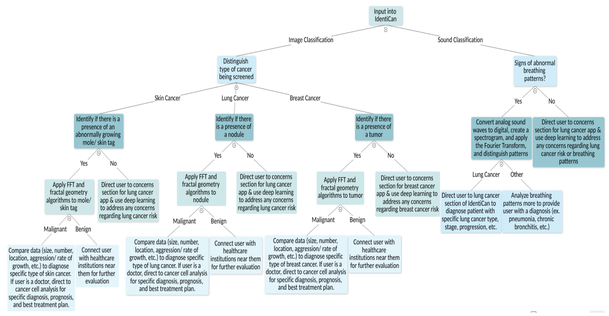
Date	Trial Number	Number of Images Used to Train Machine	Benign or Malignant Outcome	Number of Iterations	Accuracy
September 4, 2022	1	5	Normal	0	0%
September 5, 2022	2	15	Malignant	3	1.60%
September 6, 2022	3	40	Benign	8	2.80%
September 7, 2022	4	135	Unclassified	9	4.70%
September 8, 2022	5	215	Benign	25	6.30%
September 10, 2022	6	370	Malignant	31	7.15%
September 11, 2022	7	465	Unclassified	40	7.85%
September 12, 2022	8	535	Normal	46	8.40%
September 13, 2022	9	725	Benign	69	10.03%
September 14, 2022	10	850	Malignant	73	11.12%
September 15, 2022	11	925	Benign	103	12.22%
September 16, 2022	12	1020	Unclassified	96	13.31%
September 17, 2022	13	1205	Benign	121	14.41%
September 18, 2022	14	1260	Malignant	180	15.50%
September 20, 2022	15	1350	Normal	227	16.60%
September 21, 2022	16	1385	Benign	218	17.69%
September 23, 2022	17	1390	Malignant	264	18.78%
September 25, 2022	18	1420	Unclassified	310	19.88%
September 26, 2022	19	1450	Benign	332	20.97%
September 28, 2022	20	1560	Malignant	362	22.07%
September 29, 2022	21	1580	Malignant	387	23.16%
September 30, 2022	22	1610	Benign	409	24.26%
October 1, 2022	23	1700	Malignant	413	25.35%
October 4, 2022	24	1720	Benign	426	26.44%
October 5, 2022	25	1745	Benign	439	27.54%
October 6, 2022	26	1750	Unclassified	443	28.63%
October 8, 2022	27	1765	Malignant	490	29.73%
October 9, 2022	28	1800	Benign	502	30.82%
October 11, 2022	29	1900	Malignant	510	31.92%
October 12, 2022	30	2010	Malignant	538	33.01%
October 13, 2022	31	2060	Malignant	540	34.10%
October 15, 2022	32	2150	Benign	573	35.20%
October 16, 2022	33	2230	Benign	602	36.29%
October 18, 2022	34	2245	Malignant	611	37.39%
October 19, 2022	35	2450	Benign	629	38.48%
October 21, 2022	36	2565	Unclassified	632	39.58%
October 22, 2022	37	2705	Benign	647	40.67%
October 24, 2022	38	2790	Benign	662	41.76%
October 25, 2022	39	2850	Normal	670	42.86%
October 27, 2022	40	2960	Malignant	684	43.95%
October 28, 2022	41	3050	Benign	703	45.05%
October 30, 2022	42	3200	Malignant	710	46.14%
November 1, 2022	43	3265	Unclassified	723	47.24%
November 3, 2022	44	3455	Benign	735	48.33%
November 4, 2022	45	3605	Unclassified	748	49.42%
November 6, 2022	46	3680	Malignant	760	50.52%
November 7, 2022	47	3725	Normal	774	51.61%
November 9, 2022	48	3800	Benign	785	52.71%
November 10, 2022	49	3950	Malignant	823	53.80%
November 12, 2022	50	4100	Benign	841	54.90%

Date	Trial Number	Number of Images Used to Train Machine	Benign or Malignant Outcome	Number of Iterations	Accuracy
September 4, 2022	1	5	Unclassified	0	0%
September 5, 2022	2	25	Squamous Cell Carcinoma	2	1.32%
September 6, 2022	3	90	Unclassified	6	2.71%
September 7, 2022	4	145	Benign	9	5.52%
September 8, 2022	5	285	Benign	14	5.90%
September 10, 2022	6	370	Malignant	24	7.84%
September 11, 2022	7	425	Melanoma	48	7.95%
September 12, 2022	8	605	Malignant	28	8.47%
September 13, 2022	9	735	Unclassified	52	10.12%
September 14, 2022	10	920	Basal Cell Carcinoma	95	11.72%
September 15, 2022	11	1085	Malignant	83	13.49%
September 16, 2022	12	1195	Benign	107	15.16%
September 17, 2022	13	1305	Unclassified	118	16.32%
September 18, 2022	14	1495	Malignant	123	21.03%
September 20, 2022	15	1560	Malignant	100	15.39%
September 21, 2022	16	1615	Squamous Cell Carcinoma	125	22.98%
September 23, 2022	17	1645	Malignant	135	25.36%
September 25, 2022	18	1785	Benign	150	27.14%
September 26, 2022	19	1800	Melanoma	152	28.62%
September 28, 2022	20	1820	Benign	173	30.06%
September 29, 2022	21	1895	Unclassified	245	32.40%
September 30, 2022	22	1915	Melanoma	310	34.66%
October 1, 2022	23	1970	Malignant	328	36.79%
October 4, 2022	24	2045	Benign	379	39.27%
October 5, 2022	25	2175	Unclassified	472	42.65%
October 6, 2022	26	2290	Unclassified	506	44.58%
October 8, 2022	27	2355	Benign	432	40.32%
October 9, 2022	28	2485	Malignant	468	42.74%
October 11, 2022	29	2530	Basal Cell Carcinoma	506	44.82%
October 12, 2022	30	2540	Malignant	510	45.05%
October 13, 2022	31	2610	Benign	546	46.59%
October 15, 2022	32	2730	Benign	572	47.34%
October 16, 2022	33	2785	Melanoma	591	48.13%
October 18, 2022	34	2845	Benign	603	49.64%
October 19, 2022	35	2960	Unclassified	618	50.37%
October 21, 2022	36	3030	Basal Cell Carcinoma	624	51.09%
October 22, 2022	37	3105	Benign	693	52%
October 24, 2022	38	3190	Malignant	647	47.14%
October 25, 2022	39	3285	Malignant	610	45.78%
October 27, 2022	40	3305	Benign	708	48.53%
October 28, 2022	41	3455	Squamous Cell Carcinoma	716	48.96%
October 30, 2022	42	3510	Malignant	732	50.14%
November 1, 2022	43	3585	Benign	748	51.04%
November 3, 2022	44	3670	Benign	730	49.26%
November 4, 2022	45	3805	Benign	779	52.66%
November 6, 2022	46	3820	Unclassified	773	51.74%
November 7, 2022	47	3895	Squamous Cell Carcinoma	802	53.92%
November 9, 2022	48	3930	Benign	809	54.86%
November 10, 2022	49	3995	Melanoma	817	55.03%
November 14, 2022	50	4010	Malignant	832	57.83%

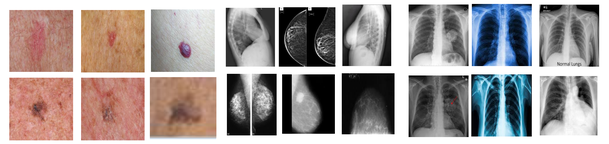


All images and graphics were created by the researcher unless otherwise noted.

Machine Learning



Sample Training Data



Fractal Geometry

- IdentiCan has fractal geometry algorithms that look for patterns in the dimensions and outlines of tumors or moles that identify whether the image is showing malignant or benign.
- Can be applied for the future prediction of growth for the future tumor or mole growth which allows for the creation of more specificized and successful treatment plans.

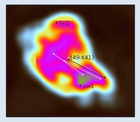


Image from www.skinvision.com [SkinVision]

Fourier Transform

- The fast Fourier Transform allows IdentiCan to able to track the complexity outline of a tumor through these algorithms which help show whether a tumor is benign or malignant.
- The distance of the tumor outline from a certain point can be quantitively represented in a wavelength format, and FFT analysis can then be applied for diagnosis.
- In sound classification, IdentiCan converts the analog sound waves to digital, creates spectrograms, and then applies the Fourier Transform to get individual frequencies, which can then be analyzed.

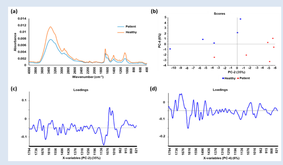


Image from www.mdpi.com [MDPI Journals]

Conclusions

- IdentiCan is successfully able to diagnose breast, lung, or skin cancer and provide a specialized treatment plan with an overall accuracy of 94.3%.
- Our app helps advance and prove research that has been done on the correlation between machine learning and cancer, in both cancer detection and diagnosis, as well as successful treatment plans.
- Our projected users are doctors and people who wish to screen themselves for cancer using an accessible and reliable platform.
- From feedback received from experts and professionals in the field, we have noted that our app fits legal guidelines and have included a page for terms of service.