Low Cost Light Source Identification in Real World Settings

Tushar Routh, Nurani Saoda, Md Fazlay Rabbi Masum Billah, Bradford Campbell
University of Virginia, Charlottesville, Virginia, USA
{tkr2w,saoda,masum,bradjc}@virginia.edu

***Abstract*—Recent studies have shown that, experiencing the appropriate lighting environment in our day-to-day life is paramount, as different types of light sources impact our mental and physical health in many ways. Researchers have interconnected daylong exposure of natural and artificial lights with circadian health, sleep and productivity. That is why having a generalized system to monitor human light exposure and recommending lighting adjustments can be instrumental for maintaining a healthy lifestyle. At present methods for collecting daylong light exposure information and source identification contain certain limitations. Sensing devices are expensive and power consuming and methods of classifications are either inaccurate or possesses certain limitations. In addition, identifying the source of exposure is challenging for a couple of reasons. For example, spectral based classification can be inaccurate, as different sources share common spectral bands or same source can exhibit variation in spectrum. Also irregularities of sensed information in real world makes scenario complex for source identification. In this work, we are presenting a Low Power BLE enabled Color Sensing Board (LPCSB) for sensing background light parameters. Later, utilizing Machine learning and Neural Network based architectures, we try to pinpoint the prime source in the surrounding among four dissimilar types: Incandescent, LED, CFL and Sunlight. Our experimentation includes 27 distinct bulbs and sunlight data in various weather/time of the day/spaces. After tuning classifiers, we have investigated best parameter settings for indoor deployment and also analyzed robustness of each classifier in several imperfect situations. As observed performance degraded significantly after real world deployment, we include synthetic time series examples and filtered data in the training set for boosting accuracy. Result shows that our best model can detect the primary light source type in the surroundings with accuracy up to 99.30% in familiar and up to 90.25% in unfamiliar real world settings with enlarged training set, which is much elevated than earlier endeavors.**

***Index Terms*—Light Source Classifying, Low power sensing**

I. Introduction

The role of lighting to human beings is not merely limited to illumination, but also impacts a person physiologically and psychologically [1], [2]. As a diurnal species, the periodicity of light exposure throughout the entire day is crucial [3]. Researchers have examined the influence of light exposure on human during different cycles of a day by studying heart rate, cortisol, concentrated body temperature (CBT), fatigue, and sleeping behavior [4]. Exacerbation of behavioral disturbances and the disrupted circadian sleep patterns have been observed in people with dementia due to improper lighting scenarios [5]. Anomalies like inadequacy or non-periodicity in melatonin production, an event that is coupled with daylong light exposure, has been found as one of the major offenders for sleep disorders that affects 50 to 70 million adults and one third of the senior population in the US [6], [7]. Not only lighting parameters, but also lighting type, especially at night, can suppress and delay the normal operation of a person's biological clock [8]. For example, avoiding blue enriched sources (most present day LEDs) is recommended by health professionals after sundown hours for quality sleeping [9]. Studies also show that careful lighting design can improve healthiness among senior citizens, Alzheimer's disease and related dementia (ADRD) patients, and others [10]–[12]. Therefore, continuous monitoring of various types of light exposure data throughout a day is imperative, particularly at nursing homes and hospitals, where lighting schemes are purposefully decorated for ensuring ambience and as a part of treatment [13].

Commonly deployed devices for sensing light contain certain limitations. Acquiring light exposure statistics during a whole day can be expensive, memory-intensive, highly power-consuming and on top of that, sensors are mostly designed to be wearable which adversely effects level of comfort and ergonomics. Even when light sensors are deployed as an immobile device, detection accuracy can vary based on sensor placement, parameter selection, adopted classification model and nature of classifiers' training set. Unfortunately, present studies cannot answer which classifier and what parameters are best suited for indoor light classification.

Adopted classifiers till to date are trained only with standalone sources, with limited examples and setups remained non interrupted throughout data collection. However, in real world, identifying environments can deviate from ideal scenario in multiple ways. Modern day lighting architectures are not isolated, rather have become dynamic and personalized through blending sources of multiple types and specific features, which creates a complex environment for specifying the major contributing source. Classifiers that are trained with limited examples will fail to identify source that lies outside the training set. Also in reality, signal patterns can randomly fluctuate during on-off/presence of noise around sources or undesirable/unavoidable interruption during acquisition, like obstruction between the source and the sensor due to human movement. As classifiers are not familiarized with such signal patterns, they tend to mis-classify at those adversaries and accuracy falls below satisfactory level. Therefore, modification of training data is a pre-requisite to get our classifiers

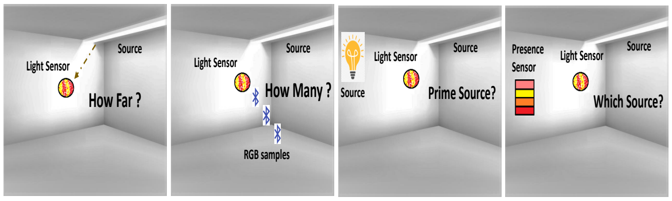


Fig. 1: An illustration of a few analysis done in this paper. From left to right a) Sensor placement, b) No of samples for classification, c) Detecting prime source in multi-source environment, d) Identifying light type at smart environments

acquainted and correctly classify sources in real world settings.

In this work, we have designed and developed a Bluetooth Low Power (BLE) enabled color sensing board for acquiring light exposure information for extended period and demonstrated how recording only RGB information can be fruitful identifying major source exposure at various times. This smart device exploits very low memory, suitable for indoor deployment and flexible to be shaped into wearable format if required. Based on sensed information, it can calculate on-the-spot lighting parameters like Lux Intensity (LI) and Correlated Color Temperature (CCT), as well as provide data to distinguish major source in background off-board. After placing this sensor at indoor atmospheres, we investigate the ideal distance from light source for placing sensors and no of samples needed, along with their dimensions for optimal classifying performance. To find the best performing classifier, we have studied multiple Machine Learning and Neural Network based classifying methods and made comparative analysis of accuracy of those classifiers in ideal/non-ideal backgrounds like multi-source/noisy/smart environments. Finally, we introduce application of Time Series Generative Adversarial Network (TimeGaN) to generate synthetic examples to familiarize light sources outside training set. We have also designed and implemented various length filters for recognizing sources from irregular signal patterns. After introducing those methods, we have observed elevated source identification accuracy in real world setup. This study will be advantageous regarding indoor deployment of light sensors for optimal performance, as well as robust, power efficient and persistent identification of source exposure in real world.

II. Related Work

Source classification techniques till date has been primarily relied on spectrum data from mini-spectrometers. *C12666MA* mini-spectrometer from Hamamatsu electronics has been the most favored which costs around $400, operates on 4.75-5.25 V range and consumes power around 30mW. A similar of its kind, HPCS300P Mini Spectrometer (price around $500) uses USB interface (500mA/5V-900mA/5V) to operate. Even, with low cost lower resolution version, like *TINYSA* analyzer (only $49), runs on battery allowing only 2/4 hours for portable use. Mini-spectrometers from *Pasco* can operate on wireless mode, but again costly (around $450) for multi-location mass deployment. Utilizing *C12666MA* mini-spectrometer, non-visual impacts of light exposure was studied in modern homes by identifying daily source exposure and recording daily sleeping hours [14]. *Spectrace*, a wearable sensor for spectrally-resolved personal light monitoring system was built to recover a diversity of spectra at different bandwidths consisting accelerometer and gyroscope to provide feedback of the current light exposure [15]. Proposed sensor was claimed to be low cost, small size to provide a high-accuracy result of spectrum-specific light intensity [16]. Low-cost and portable spectrometer using CMOS-based sensors was designed which is able to detect wavelengths in a range from visible to NIR region. Named *AvaSpec-Mini2048CL spectrometer*, different types of electric lights, along with natural light source were chosen for capturing class variation and MLP model was used for data reconstruction. Prediction errors were calculated for different indoor and outdoor conditions after comparing with *Wavego* [17]. Fernandez [18] utilized RGB information from *TCS3414CS color sensor* and *ADJDS311 color sensor* to classify various artificial sources (34 LED, 16 incandescent and 6 fluorescent sources) and selecting a model estimation of Color Rendering Index (CRI) and Correlated Color Temperature (CCT). Ma, Bader and Oelman [19] did similar kind of research with *TSL2561, ISL29125* color sensors, *AM1815CA, POW11D2P* solar cells and *USB2000+* spectrometer, where sensor data for Halogen, Fluorescent, LED and Incandescent bulbs were collected via USB interface and I-V tracers and KNN, SVM and Decision tree algorithms were utilised for classification for the most part. It has been displayed that even with higher intensity interference from other sources, ML based approach can typify sources with only 62.5% outside training specimens.

III. Limitations of Earlier Approaches

Indoor light characterization with spectrum analyzer is high-priced and data acquisition process is intensely energy and memory hungry. For real world deployment, cumulative energy expenditure becomes significant for daylong operation. Moreover, higher spectral resolution data throughout the day may conglomerate that appeals humongous memory stack. When our goal is metering the source type with common lighting parameters, high resolution spectral information is not quintessential.

Patients/senior citizens who have limited movement, carrying device for the whole day with other appliances may offer discomfort and undesirable for gathering lighting aspects at indoors. Where person spend most of his/her hours under the roof, easy to install indoor smart sensors can uncover their round the clock lighting exposure. IoT based flexible RGB sensors should come into play. In addition to offering wearability/mobility, they can be deployed as a fixed room light sensor for accessing lighting information from practical distances at low energy cost and operate for an extended period without power/memory replacement.

When light sensors are carried by human, relative distance between the source and sensor position is unmanageable. But in case of indoor deployment, placements can be climacteric for elevated performance, which unfortunately, were not dissected in earlier investigations.

Whether remotely installed or designed as a wrist-band device, intelligence regarding number of samples and their sizes are critical for classifiers to determine the source type, where methodical studies are few and far between.

With spectrum analyzer, magnitude at a particular wavelength and transient waveform shape were adopted for typifying, which is inaccurate, as sources of different types share common spectral range. Magnitude of the sensed parameters can fluctuate based on the relative distance between the source and the sensor and at last, data acquisition in real life cannot be always conducted only during transient switch on/off phases.

When machine learning algorithms were called into play, the training size chosen for experiments were too small to draw any conclusion. Sometimes same sources with varying intensities were trained with, which fails to encompass any substantial portion of most common and evolved varieties of bulbs available at market. Even Sunlight possess indoor variations throughout the time of the day, weather and surroundings, which can be misidentified if not acknowledged. As a consequence, classifiers report near perfect accuracy with familiar sources, but performs poorly after encountering sources outside training set. Neural Networks (NN) along with Time series based analysis should come into play for their reputation in pattern recognition for unseen examples and seen examples in non ideal scenarios.

Previous approaches of categorization were mostly based on considering isolated single sources, whereas captured readings may get influenced from another source or from any RGB element's presence near the sensor that can misdirect classifier towards pinpointing wrong class. In modern indoor lighting, sources keeps on/off based on person's presence. This can completely shatter the steady state RGB pattern and makes classification task complicated.

IV. Properties for classification and effects round the clock

In this work, we have considered three most widely used indoor bulbs with sunlight: Incandescent lamps, Compact Fluorescent Light bulbs (CFLs) and Light Emitting Diodes (LEDs)(figure 2. Radiation is generated through heating tungsten filament for incandescent bulbs. CFL mostly offers "cool white light" and spectrum exhibits certain spikes during the startup phase [20]. Led delivers radiance over a wide band of wavelengths, like soft white (2700K-3000K), cool white (3100K-4000K), daylight (5000K-6000K) etc. Emissive surfaces of LEDs are highly-concentrated, illuminance of which can be 1000 times higher than recommended level [21]. Although sunlight covers the broadest spectrum, its nature is dynamic, intensity and color components of light (wavelengths) change with the time of day, time of year, the weather and the location on earth.

CFLs and LEDs may be energy efficient but emit more unhealthy blue light that disrupts triggering the release of the biological stimulation [21]. Absorption of blue light component changes with age and increases with light intensity. Bright Sunlight is the most powerful source for blue enriched light (upto 1,500 μW/cm^2) can boost maintaining healthy sleeping order, whereas LED computer screen with blue illuminance around 30 μW/cm^2 couple of hours before bed can promote lower melatonin secretion [22]. That is why maintaining daylong healthy light receptiveness through appropriate class of light is imperative.

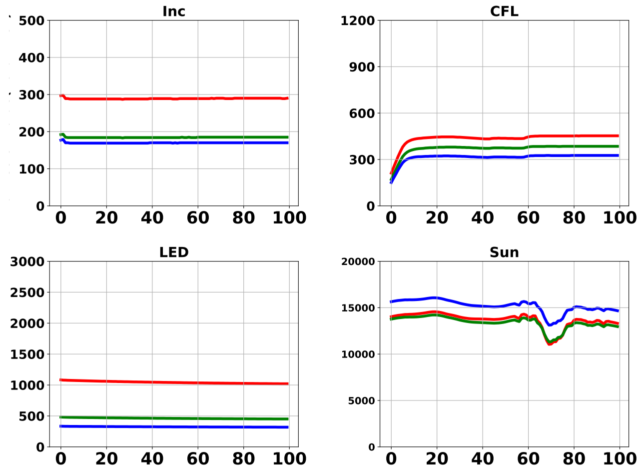


Fig. 2: **100 samples of sensed RGB values for each type of light source (x-axis:sample no., y-axis: hex value): Incandescent ("soft white", 40 W), CFL ("natural daylight", 13 W), Led ("soft white", 9 W), Sunlight (open, 12:45 pm)**

V. Methodology

To address power efficient elongated operation of light sensing, we develop a low power color sensing board (LPCSB) dedicated to sense lighting information from the near around environment and transmit data in a wireless/local fashion. Although there are multiple color sensors in the market, most of them are not cost effective, are large dimensional, energy inefficient and require to relay information to central hub for further analysis mostly through wired connections. Our goal was to develop small scale, low-cost, mobile, lightweight task specific sensor, that is unobtrusive to already installed systems in that surroundings and easy to deploy as smart room sensor or as wearable systems in future. LPCSB advertises BLE packets containing RGB, clear value (related to intensity), color temperature and lux information of a light source (calculated from RGB values), which enables user to place the board in inaccessible/unreachable areas, connect with BLE receivers and then deliver sensed values as instructed. Moreover, the system consumes extremely low power, as a result the power source does not need to get replaced often which lowers down the maintenance hazards. In addition, information can be captured from a distance and analysed in any platform of users choice (for example, smart watch or remote servers). For classification, we use this board only as an advertiser to advertise a BLE data packet containing ID, raw data (clear, red, green, and blue) split into two bytes per color, color temperature and lux of the measured light calculated from raw rgb values and the number of the latest packet being advertised.

Raw data measurements taken from the color sensor reveal the amount of red, green, and blue components that compose the unfiltered light. Utilizing RGB info, we calculate and advertise LI and CCT of the measured light value from the sensor [23].

We then collect data from 27 different bulbs (9 from each of LED, Incandescent and CFL) for acquiring both inter-class and intra-class variation of RGB values. To achieve true nature of each light by minimizing influence from other sources, we decide to carry out all the measurements (for artificial bulbs) in a dark room. For collecting sunlight data, we expose sensor to sun in diverse conditions and scenarios, which includes taking data from sunrise to sunset, during heavy rainy, foggy and drizzling days. Inconsistency of sunlight RGB information may also derive from contrasting indoor conditions (location, window glass material, with and without blinds etc.). To accommodate them into our training set, we collect sunlight data in different buildings and also in various corners of a building.

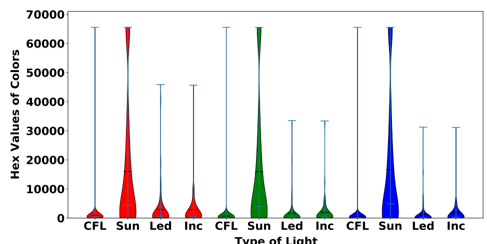


Fig. 3: **Violin plot depicts RGB values of different classes of light source lie in common range which makes it difficult to categorize based on cutoff intensity (Mean and Median values are marked with white and black lines)**

In practice, electromagnetic light waves experience reflections from nearby structures. Finally sensed complex signal, deriving from contrasting scenarios and mixed with direct and indirect components, simply do not follow inverse square law of radiation and generates irregularity in RGB values. To acknowledge magnitude variability and irregularity of RGB features based of sensor placement, we capture artificial light data at five different distances.

For analysis, we collect 500 samples for each observation. To determine optimal sample size for classification, we divide our collection window size from 10 samples up to 125 samples. Figure 3 shows RGB distribution of all the sources dealt in this study. As discovered, unalike sources share common RGB spectra and magnitude, which turns it problematic to differentiate solely based on RGB threshold. t-SNE visualization of RGB values also reveals the fact that dissimilar light sources are linearly inseparable (figure 4). That's why we investigate multiple Machine Learning(ML) and Neural Network(NN) algorithms to distinguish each type of source (shown in Table I). ML and NN models are independent of feature magnitude after scaling and capable of non linear classification. As RGB signals contain resemblance with image data (both are primarily three channel information),

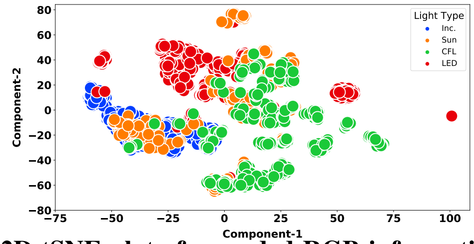


Fig. 4: **2D tSNE plot of recorded RGB information reveals linearly inseparability of source clusters**

TABLE I: Methods of classification

Method	**Acronym**	**Tuned Par./ Model Par.**
Decision Tree	DT	*criterion, max depth*
Random Forest	RF	*max depth, max features, min samples leaf, min samples split, n estimators*
Gaussian Boost	GB	*learning rate, max depth, min samples leaf, min samples split, n estimators*
Naive Bias	NB	*var smoothing*
K Nearest Neighbor	KNN	*metric, n neighbors,weights*
Logistic Regression	LR	*C parameter, penalty*
Support Vector Machine	RBF (SVM-Rad) Linear (SVM-Lin) Polynomial (SVM-Poly)	*C parameter, gamma*
Multilayer Perceptron	FNN	*No of layers:7, Dropout:20%, Activation:relu, softmax, Optimizer=SGD, Loss = Categorical cross-entropy*
Convolutional Neural Network	CNN-1D CNN-2D	*No of layers: 6 (1-D)/7 (2-D) , No. of filters: 64/32(1-D),64/32/16 (2-D), Kernel Size = 2 × 2 (1-D), 3 × 3 (2-D), Padding=same, optimizer= Adam, Dropout:20%*
Long Short Term Memory	LSTM	*No of layers: 4 , output dimension= 50,optimizer= Adam*

we inspect both feedforward Multilayer Perceptron Models (MLP) and Convolutional Neural Networks, with 1-D and 2-D filters (CNNs) for categorization. As sensed data is time series based, we have also included Long Short-Term Memory (LSTM) network for typifying.

After training and fine tuning our chosen classifiers with controlled environment data, we record performance of each classifier based on different size sample window and sensor placement. While training, we have scaled, normalized and divided the balanced dataset into training, test and validation sets (80%, 10% and 10% respectively). For better evaluation and to ensure representation from each group, we have implied stratified 10 fold cross validation by tuning classifiers to their best hyper parameters using *Gridsearch*.

For identifying primary source in a multi-source environment, we have blended RGB values light sources and observe whether our classifiers can identify the primary source. While mixing, we have made sure that the RGB values from second/interfering source never goes past values from

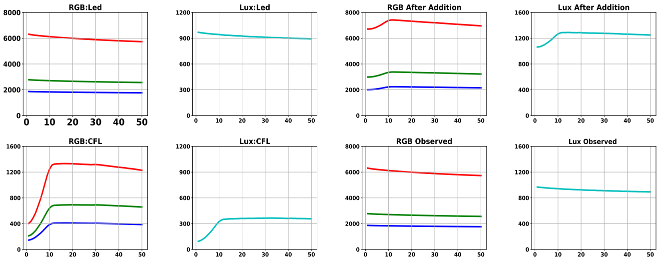


Fig. 5: **RGB observations (x-axis: sample no., y-axis: raw RGB values) along with lux intensities (x-axis: sample no., y-axis: lux/m^2) from multiple sources, LED was set as primary and CFL as secondary source. Observed outputs deviate from simple addition**

primary source, as classifier is expected to determine the major contributor between/among sources. In previous work like [19], only constructive interference has been considered as the consequence of overlapping. For further investigation, we place two light sources near the sensor and compare the resultant with simple theoretical addition. We find that they differ by a large margin, both in RGB and in lux domains (figure 5).
Based on the phase difference resulting from positioning of both sources at sensor point, a numerous blending ratio is possible. Variety of mixture represents blending of constant positioned sources at different sensor placement or positioning of sources at different locations, sensed at the same spot. However, highest possible deviations are amalgamation of identical and opposite phases. As our goal was to testify our classifiers, we have only added those extreme cases that can lead into inaccuracy with the highest probability.

Now we deploy our sensor in real life testbeds and record performances. Based on our investigation, we have seen that even after training the best model with fine tuning and wide ranging examples, performance has deteriorated substantially, especially in recognizing unfamiliar artificial sources in some observations or at unexpected events like source transitions and human movements.

With limited amount of data, machine learning models tend to over-fit and become problematic for non-linear classification. However, at the same time, it is unrealistic to include all the light source available in the market in our training set. To surpass this limitation, we have generated equal number of synthetic examples of captured data and based on current data distribution by utilizing TimeGAN. GAN generated time series RGB examples can generate realistic data for superior segregation of different light classes by adding excluded examples from source distribution. Light source classifiers is then expected to perform with higher accuracy in an environment containing unfamiliar sources. Figure 6 demonstrates distribution of first two principal components of real and synthetic examples generated using TimeGAN.

Fluctuation during data acquisition may occur arbitrarily and for unknown duration, where sensors may receive transient rather than imminent information. When our sensors records zero RGB values, it is practically impossible to detect the source type. But if it senses non-zero values even for some

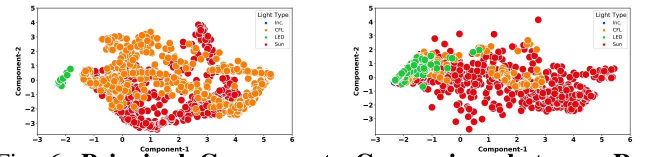


Fig. 6: **Principal Components Comparison between Real (left) and Generated Synthetic Examples (right) using TimeGAN**

duration, we may utilize that information for source classification. To familiarize our classifier models with those events, we have designed filters of different window sizes and randomly implemented them within acquisition timeframe (shown in figure 7). These examples are also expected to familiarize our classifiers with scenarios where switch on/off is non-periodic or presence of sudden obstacles in between source and sensor.

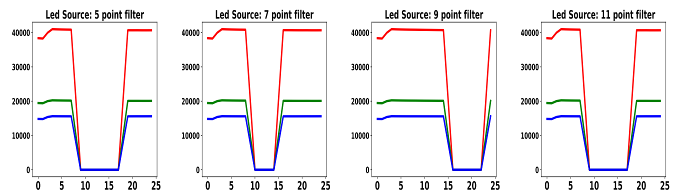


Fig. 7: **Applying filters of different sizes on a LED source to capture fluctuations in a 25 RGB sample window**

To the best of our knowledge, TimeGAN method along with filter designing have been implemented for the first time for light source classifying. After including both filtered and synthetics examples in our training set, a comparison between classification accuracy has been presented between limited and extended training set.

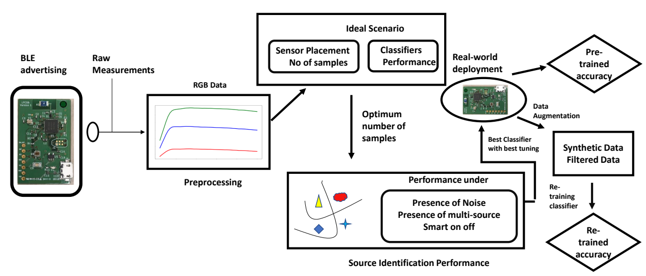


Fig. 8: **An overview of the workflow**

VI. Implementing LPCSB

LPCSB is a printed circuit board (PCB) that interfaces *TCS3475* sensor and is regulated with *nRF51822* micro controller. It is qualified to communicate over 2.4GHz Bluetooth Low Energy (BLE), flexible enough to operate in two way (transceiver mode) or one way (advertising) mode, as needed. System has a dimension of roughly $24mm \times 39.5mm$, suited to get fit and comfort as wearable devices. For low power consumption and simplification, *nRF51822* micro controller components were limited to only clock circuits, 3.3 V regulatory circuitry and power supply connector in the final design. *Micro Reach Xtend (FR05-S1-N-0-110) Chip Antenna* was assembled to establish communication and fit in PCB, plus USB connector for supplying power. Fully assembled LPCSB

Fig. 9: **Fully assembled LPCSB**

TABLE II: Approximate cost of major LPCSB components

Component	**Price in Bulk**
20 board/panel 2-layer standard thickness PCB	*$3.8*
NRF51822QF Bluetooth® 4.0/2.4 GHz RF SoC	*$3.0*
TCS3472 RGB + Clear Color Sensor	*$2.0*
MicroReach (FR05-S1-N-0-110) Chip Antenna	*$1.8*
EPSON-FA-128 (MHz RANGE CRYSTAL UNIT)	*$0.40*
MAX887EZK33+T 3.3V Linear Regulator	*$0.40*
BAL-NRF01D3 Transformer Balun	*$0.20*

can be seen in Figure 9.

The power regulation section consisted of a micro-USB B-type connector, a green LED indicator circuit, MAX887EZK33+T Low-Dropout 300mA 3.3V Linear Regulator, and several bypass capacitors meant to help stabilize the input / output voltage and current in case of supply fluctuations. With the help of *BAL-NRF01D3 transformer balun* for impedance matching and "LightBlue" phone app for monitoring, we have tested BLE radio transmitter inside *nRF51822*. Using the "nrf5x-base" and "Adafruit-TCS34725" GitHub repositories as design references, we have instructed the *TCS34725* through *nRF51822* to measure the ambient light and send the resulting values. Red-filtered, green-filtered, blue-filtered, and clear (unfiltered) diodes data of *TCS34725* sensor is stored as a 16-bit value, split between two registers. We have further calculated the color temperature of the light in degrees Kelvin and the lux in lumens per square meter, using formula provided by Adafruit. Figure 10 represents energy intake per cycle of LPCSB, where sensor reading is followed by a BLE advertisement event. If we set parameters to classify source within a minute, avg current drawn is per sampling is around 0.22mA and the system can operate upto 45 days with conventional 3.3V Lithium batteries without replacement. For mass deployment, as shown in Table II, LPCSB is notably cheaper than mini spectrometers .

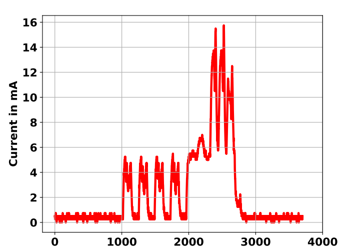


Fig. 10: **LPCSB with sensing and advertisement events**

VII. Evaluation

We have analyzed prediction accuracy in different background and have recorded mean values of classifying accuracy (with standard deviations). As initial accuracy were high with only using RGB data, we have discarded clear value, lux intensity or color temperature readings for classification. Although artificial lighting landscapes do not change very often at indoor, classifiers should be robust enough there to classify under inexperienced screenplays with factual mishaps. We start with the ideal scenarios to fix parameters for indoor deployment and then progress with evaluating non-ideal incidents with the settled values. We have illustrated the findings mostly with violin plots, which depict distributions of numeric data through density curves on the both sides of the mean value. Accuracy values exceeding 100% in those plots were trimmed.

A. Prediction in known scenario

Upper plot of Figure 11 illustrates variability of mean accuracy for different classification techniques with varying number of samples. Observation reveals accuracy is not linearly proportional with number of observed RGB samples. The best average result is achieved with 50 samples, although average accuracy with 10 and 25 samples were also near 90%. Lower

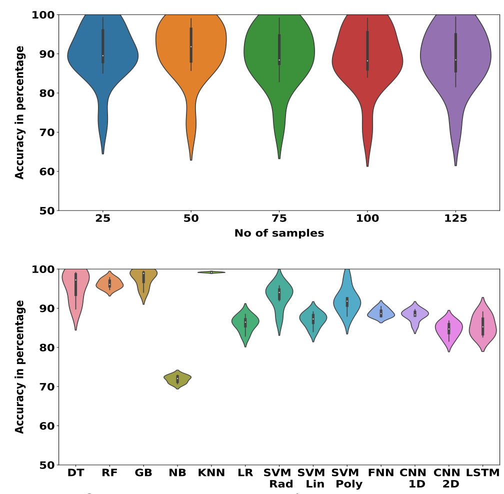


Fig. 11: **Performance comparison among ML and NN methods for LPCSB at indoor where training windows were varied from 10 to 125 samples: 50 samples triggers the best performance and KNN was the best performer**

plot of Figure 11 represents a comparative analysis among classifiers, trained with different sample sizes. It uncovers that overall performance of ML algorithms is better than NNs at ideal and known scenario. K-Nearest Neighbor (KNN) triumphs for generating highest and most consistent accuracy among all.

To carry on with KNN, our goal is to find the sweet spot for balancing number of samples with accuracy. After observing KNN accuracy with varying samples window size, we conclude 25 samples window exhibits the combination of highest accuracy and lowest standard deviation. We continue our analysis with 25 samples window for both capturing the transient and stable state of radiation and accommodating minimum number of samples at packet loss scenarios.

B. Placement of sensor

Our goal was to observe if we had the liberty of deploying sensor at any distance from the source indoor, where should we place it for yielding maximum identifying accuracy. We place LPCSB at 5 different distances for Inc, CFL and LED bulbs, starting from 50 cm to 150 cm to observe whether placement of sensor plays any role in classifiers' performances. Our analysis shows placing sensor at 100 cm can detect the background source with maximum accuracy (figure 12).

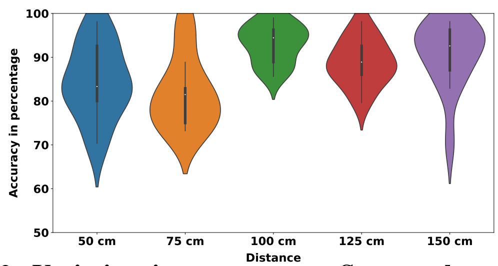


Fig. 12: **Variation in accuracy vs Sensor placement distance: 100cm yields the maximum mean accuracy**

Now, we focus on non-ideal situations with known sources/scenarios and testify classifiers performance by setting up 25 samples window length. For investigation, as before, we include 80% examples in our training set to familiarize our classifier and 10% each for validation and test sets.

C. Typifying in multiple source environment

To fabricate multi source environment, RGB values from second source was mixed at different amount, varying from 20% to 80%, of the intensity of the primary source. Highest possible deviations were included for analysis (through addition and subtraction of RGB values of primary and secondary sources) and mixing signals from all possible combinations (LED/CFL, LED/Sunlight, Sunlight/LED+CFL etc.). Our study reveals although classifiers accuracy declines with increasing mixture ratio but overall performance do not fall significantly in multi source environments (figure 13).

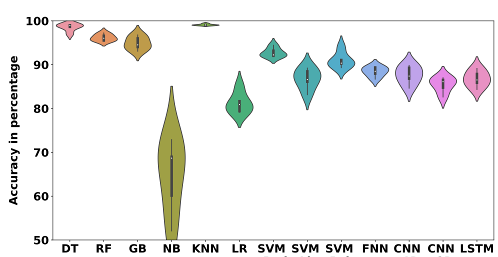


Fig. 13: **Overall accuracy at multi source environments where secondary source intensities were varied in between 20% to 80% of the primary**

D. Identifying in presence of random noise

In real world, nearby elements can act as a noise source by reflecting particular component of light which can escalate or descend the sensed values. However, by placing RGB reflecting elements nearby, we find that finally recorded value contain very were small interference. We vary the influence randomly from 0% to 5% of maximum RGB values (without noise) and enlist the performances (figure 14). As recorded,

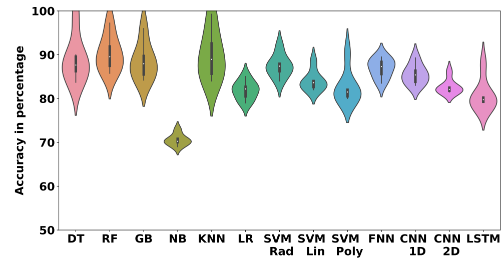


Fig. 14: **Overall accuracy based on randomly varying RGB values in between 0% to 5% of the major source intensity**

accuracy decreases with increasing intensity of perturbations. All inclusively, NN based classifiers can withstand turbulence better than ML based classifiers. Random forest performs best among ML algorithms (mean accuracy 84.46%) where accuracy score of KNN was close to that (mean accuracy 83.17%).

E. Detection Precision in smart environment

For source detection in smart environments, we vary the on/off duration of sources and also the cutoff points, from 20% to 80% of the maximum value. Final RGB signal patterns had major shifts from initial pattern based on threshold. Now

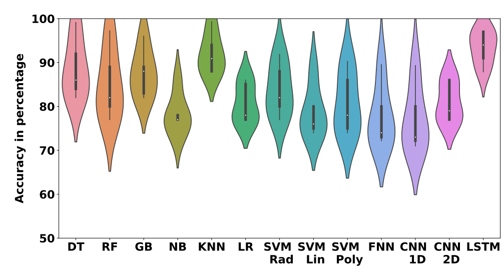


Fig. 15: **Overall accuracy in primary source detection at smart environments**

evaluating the performances to identify the altered patterns, we inspect that LSTM is the best performer (figure 15). If we enlarge the accuracy in KNN case based on threshold, we can see that accuracy of classification decreases with lowering threshold values (figure 16). So balancing threshold is a prerequirement for desired accuracy in smart environments.

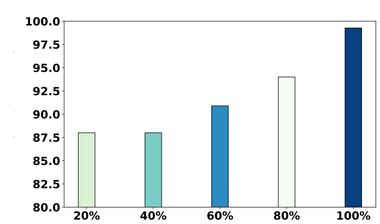


Fig. 16: **Declining detection accuracy of KNN classifier was observed with lower cutoff settings**

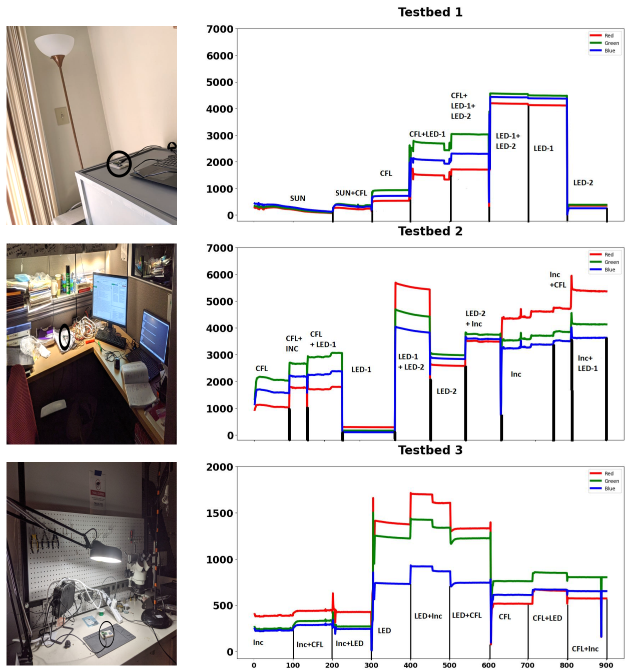


Fig. 17: **RGB signals in test-beds with accuracy, circles point placement of LPCSB (x-axis:sample no., y-axis: hex value)**

F. Real world deployment

Here, we deploy our board in real world settings. For classification, we have singled out KNN as our classifier, for exhibiting the most consistent performance in all scenarios and trained it with all ideal/non-ideal examples from controlled atmosphere. We conduct 3 experiments at 3 different test beds: (1) Household, (2) Lab environment-1 and (3) Lab environment-2. All the test were done with completely unknown artificial bulbs. Experiments included single source, mixed light source and arbitrary switching of bulbs scenarios (figure 17).

G. Analysing misidentified examples

After analysis, we observe that classification accuracy has been degraded unexpectedly. Classifier got confused during few transition events. We also observe that the faulty predictions were not common for any particular light. Moreover, for artificial lamps, a single source at different distances have been typified as different classes. While detecting RGB spectrum of sunlight during sunrise and sunset, classifier has been misguided.

Now, we retrain our KNN classifier with extended training set and re-record the accuracy. Our investigation reveals that performance of KNN classifier has been significantly improved (figure 18), especially in few cases of sudden movements and random switching between sources (figure 19). Even with elevated training, we observed examples that were failed to get correctly identified. A few of them have been listed below (figure 20). Again, no single pattern of mis-classification was discovered.

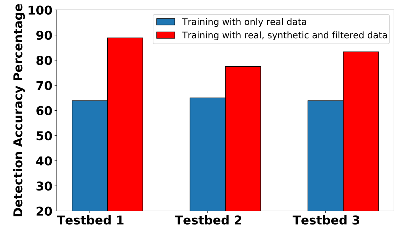


Fig. 18: **Accuracy reveals KNN with extended training set exhibits superior performance in unfamiliar environments**

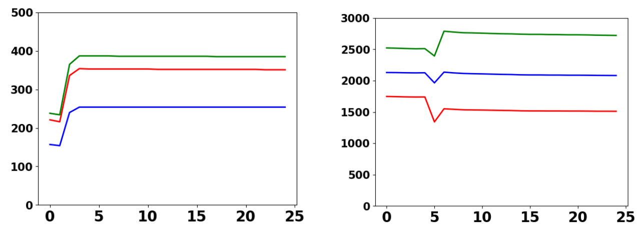
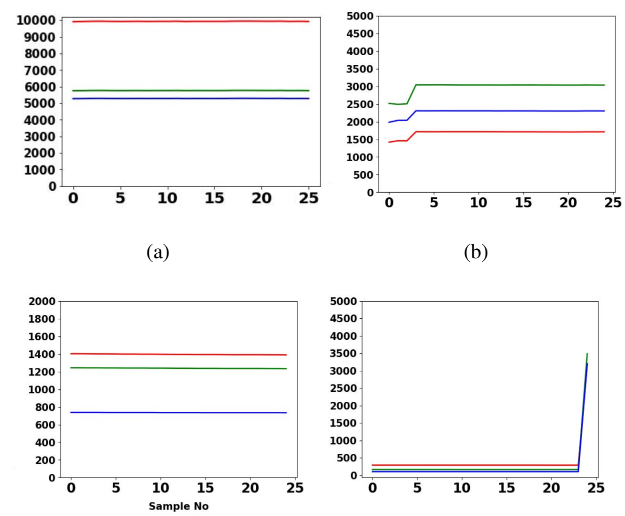


Fig. 19: **Correct Predictions with extraneous training set (x-axis: sample no., y-axis:hex value) : During Switch-over (left):Prev-Led, Now-CFL. During random movement (right):Prev-CFL, Now-Sunlight**



(a) (b)

(c) (d)

Fig. 20: **Miscategorized examples with Incandescent ("work white", 150 W),Led ("warm yellow light", 40 W) and CFL ("T9, 6400K", 22 W) bulbs (x-axis: sample no., y-axis: hex value): (a) Inc. as Sunlight (b) Led as Sunlight (c) Led as Inc. (d) CFL as Led**

VIII. Discussion

In this experiment, we have used *TCS3475* sensor with 1X ADC gain with an integration time of 700 ms, which allows us to read color values up to value 65535. Based on the place of interest for source detection, color sensor parameters like integration time and ADC gain settings can be modified for increased sensitivity at low light levels. Sensors can be preset only to record data if there is any certain amount of change in value and discard values that are below threshold. Sampling and Advertising rate can also be adjusted for better power management on the transmitting side. However, if the rate is too high and number of samples for identification are prefixed, it may fail to capture amount of variation needed for classification. On the other hand, too low sampling rate will result in unnecessary delay in classification process. Moreover, color sensors like TCS3475 has a limitation regarding integration time and highest value that can be recorded for a color channel. When running the control tests, we did not test light bulbs with colored glass in detail or rotating search lights. As BLE technology has range limitation, on board storing and processing can be beneficial to minimize packet loss but conceivably will require higher memory and processing power.

IX. Conclusion and Future work

For accurate identification of light exposure, light sensors need to encounter all the scenarios we have discussed here. What we have monitored is that a single classifier is not the best performer in all the landscapes. In addition, inaccurate identification was not bulb specific. For indoor deployment, placement of sensors, along with the number of samples considered for source identification play pivotal roles. With KNN, sensing 25 samples at a distance of 100 cm achieves accuracy up to 99.30% in constrained cases, compare to 100% in identifying indoor light among LED, CFL, Inc. and Halogen [19] and 100% in distinguishing among Warm Led, Cool Led, Halogen, CFL and solar simulator [24] (in both the cases, architectures were trained with only one example of each type with varied lux intensities). Recognition of primary source in a multi-source environment, classifiers' mean accuracy was 98.96%, compare to 100% in [19] and 85.4% in [24]. However, after adding filtered and synthetic data, our highest mean accuracy was 90.25% for unfamiliar synopses, compare to only 62.5% in [19]. Training classifier with limited data set may fall short for classification in real world setting, where adding synthetic and filtered data can elevate the performance. Finally, interference like accidental movement can hit hard the performance, so stable environment is advantageous.

Our future endeavors include readjusting sensor parameters that further minimize energy intake for daylong operation, keep high accuracy intact and help us gathering knowledge regarding primary source around the environment as quickly as possible. On board classification approach can be embraced for developing self-contained systems to store, analyze and publish outcomes as a package like smart watch. We also plan to conduct time series feature based identification which may require higher memory and power for calculation but can identify primary source in nearby atmosphere with higher accuracy than utilizing only raw RGB values. Mis-classified and adversarial examples can be included in the training set for upgrading robustness.

References

[1] R. Kralikova, M. Piňosová, and B. Hricová, "Lighting quality and its effects on productivity and human healts," *Int. J. Interdiscip. Theory Pract*, vol. 10, pp. 8–12, 2016.

[2] L. Bellia, F. Bisegna, and G. Spada, "Lighting in indoor environments: Visual and non-visual effects of light sources with different spectral power distributions," *Building and environment*, vol. 46, no. 10, pp. 1984–1992, 2011.

[3] L. Sahin, B. M. Wood, B. Plitnick, and M. G. Figueiro, "Daytime light exposure: Effects on biomarkers, measures of alertness, and performance," *Behavioural brain research*, vol. 274, pp. 176–185, 2014.

[4] M. Ruger, M. C. Gordijn, D. G. Beersma, B. de Vries, and S. Daan, "Time-of-day-dependent effects of bright light exposure on human psychophysiology: comparison of daytime and nighttime exposure," *American Journal of Physiology-regulatory, integrative and comparative physiology*, vol. 290, no. 5, pp. R1413–R1420, 2006.

[5] L. Baandrup and P. J. Jennum, "Effect of a dynamic lighting intervention on circadian rest-activity disturbances in cognitively impaired, older adults living in a nursing home: A proof-of-concept study," *Neurobiology of Sleep and Circadian Rhythms*, vol. 11, p. 100067, 2021.

[6] F. Ozkalayci, U. Kocabas, B. U. Altun, S. Pandi-Perumal, and A. Altun, "Relationship between melatonin and cardiovascular disease," *Cureus*, vol. 13, no. 1, 2021.

[7] "Sleep and sleep disorder statistics," 07 2021. [Online]. Available: https://aasm.org/

[8] L. Tähkämö, T. Partonen, and A.-K. Pesonen, "Systematic review of light exposure impact on human circadian rhythm," *Chronobiology international*, vol. 36, no. 2, pp. 151–170, 2019.

[9] C. Ticleanu, "Impacts of home lighting on human health," *Lighting Research & Technology*, vol. 53, no. 5, pp. 453–475, 2021.

[10] N. Shishegar, M. Boubekri, E. A. Stine-Morrow, and W. A. Rogers, "Tuning environmental lighting improves objective and subjective sleep quality in older adults," *Building and Environment*, vol. 204, p. 108096, 2021.

[11] K. Missildine, "Sleep and the sleep environment of older adults in acute care settings," *Journal of gerontological nursing*, vol. 34, no. 6, pp. 15–21, 2008.

[12] M. M. Ohayon and C. Milesi, "Artificial outdoor nighttime lights associate with altered sleep behavior in the american general population," *Sleep*, vol. 39, no. 6, pp. 1311–1320, 2016.

[13] M. Figueiro, M. S. Rea, E. Saldo, M. Rea, K. Kubarek, and J. Cunningham, "Developing architectural lighting designs to improve the health and well-being of older adults," *AIA report on University Research*, vol. 3, pp. 3–26, 2008.

[14] S. W. Cain, E. M. McGlashan, P. Vidafar, J. Mustafovska, S. P. Curran, X. Wang, A. Mohamed, V. Kalavally, and A. J. Phillips, "Evening home lighting adversely impacts the circadian system and sleep," *Scientific reports*, vol. 10, no. 1, pp. 1–10, 2020.

[15] F. S. Webler, G. Chinazzo, and M. Andersen, "Towards a wearable sensor for spectrally-resolved personal light monitoring," in *Journal of Physics: Conference Series*, vol. 2042, no. 1. IOP Publishing, 2021, p. 012120.

[16] H.-Y. Tsai, F.-C. Su, C.-H. Chou, Y.-H. Lin, K.-C. Huang, Y.-J. J. Yang, L.-W. Kuo, L.-D. Liao, and H.-S. Yu, "Wearable inverse light-emitting diode sensor for measuring light intensity at specific wavelengths in light therapy," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 5, pp. 1561–1574, 2019.

[17] A. Amirazar, M. Azarbayjani, M. Molavi, and M. Karami, "A low-cost and portable device for measuring spectrum of light source as a stimulus for the human's circadian system," *Energy and Buildings*, vol. 252, p. 111386, 2021.

[18] A. Fernández-Montes, L. Gonzalez-Abril, J. A. Ortega, and F. V. Morente, "A study on saving energy in artificial lighting by making smart use of wireless sensor networks and actuators," *IEEE network*, vol. 23, no. 6, pp. 16–20, 2009.

[19] X. Ma, S. Bader, and B. Oelmann, "Characterization of indoor light conditions by light source classification," *IEEE Sensors Journal*, vol. 17, no. 12, pp. 3884–3891, 2017.

[20] A. Panahi, "The health risks associated with energy efficient fluorescent, leds, and artificial lighting," in *Photonics Applications for Aviation, Aerospace, Commercial, and Harsh Environments V*, vol. 9202. International Society for Optics and Photonics, 2014, p. 92020K.

[21] K. M. Zielinska-Dabkowska, "Make lighting healthier," 2018.

[22] A.-M. Lahr, "What we know—and don't know—about blue light."

[23] A. Sarris, "LPCSB: A Device to Distinguish Between Natural and Artificial Light," Master's thesis, University of Virginia, Charlottesville, Virginia, 2020.

[24] Y. Zhang, "On the importance of light source classification in indoor light energy harvesting," 2018.