Spectral index to improve the extraction of built-up area from WorldView-2 imagery

Adeniyi Adeyemi[a,*], **Abel Ramoelo**[b,c], **Moses Cho**[d,e], **and Cecilia Masemola**[f]

[a]University of South Africa, College of Agriculture and Environmental Science, Department of Environmental Science, Pretoria, South Africa
[b]Scientific Services, South African National Parks, Pretoria, South Africa
[c]University of Pretoria, Department of Geography, Geoinformatics and Meteorology, Pretoria, South Africa
[d]Precision Agriculture, Council for Scientific and Industrial Research, Pretoria, South Africa
[e]University of Pretoria, Department of Plant and Soil Sciences, Pretoria, South Africa
[f]University of KwaZulu-Natal, Department of Agriculture, Engineering & Environmental Science, Pietermaritzburg, South Africa

Abstract. Globally, the unprecedented increase in population in many cities has led to rapid changes in urban landscape, which requires timely assessments and monitoring. Accurate determination of built-up information is vital for urban planning and environmental management. Often, the determination of the built-up area information has been dependent on field surveys, which is laborious and time-consuming. Remote sensing data are the only option for deriving spatially explicit and timely built-up area information. There are few spectral indices for built-up areas and often not accurate as they are specific to impervious material, age, colour, and thickness, especially using higher resolution images. The objective of this study is to test the utility of a new built-up extraction index (NBEI) using WorldView-2 (WV-2) to improve built-up material mapping irrespective of material type, age, and color. The new index was derived from spectral bands such as green, red edge, NIR1, and NIR2 bands that profoundly explain the variation in built-up areas on WV-2 image. The result showed that NBEI improves the extraction of built-up areas with high accuracy [area under the receiver operating characteristic curve, (AUROC) = ∼0.82] compared to the existing indices such as built-up area index (AUROC = ∼0.73), built-up spectral index (AUROC = ∼0.78), red edge/green index (AUROC = ∼0.71) and WorldView-Built-up Index (WV-BI) (AUROC = ∼0.67). The study demonstrated that the new built-up index could extract built-up areas using high-resolution images. The performance of NBEI could be attributed to the fact that it is not material-specific, and would be necessary for urban area mapping. [DOI: 10.1117/1.JRS.15.024510]

Keyword: WorldView-2; spectral indices; built-up; very high resolution.

Paper 210046 received Jan. 24, 2021; accepted for publication Apr. 8, 2021; published online Apr. 26, 2021.

1 Introduction

The global urban population proliferated from 220 million to 2.8 billion over the twentieth century.[1] This unprecedented increase in population concentration in cities led to rapid urban landscape changes.[2,3] The highest rate of urbanization and associated land use or cover changes have been observed in developing countries.[4] Over the last decades, Southern Africa has been facing significant land use and land cover changes, such as loss of natural land, i.e., forest or plantations, agricultural lands, and grasslands coupled with growing built-up impervious surfaces, which are developed and constructed artificial surfaces that water cannot infiltrate to

*Address all correspondence to Adeniyi Adeyemi, adedayoadeyemi01@gmail.com

1931-3195/2021/$28.00 © 2021 SPIE

reach the soil, such as buildings or rooftops, paved roads, driveways, roads sidewalks, parking lots, and so on.[5] The dynamic nature of land use or cover is associated with economic benefits and improved life. The latter effect has been associated with a series of ecological, environmental, and climatic issues.[1,6,7] Therefore, in recent years, detecting these human-made features has attracted increasing interest especially for understanding their adverse impact in the context of planning and improvement of the environment. Often, the determination of the built-up area information has been dependent on field data collection, which is tedious and time-consuming.

Satellite remote sensing data have served as a useful source of built-up area information in previous studies due to large-area coverage and short revisit cycles.[8] Among these studies, Landsat imagery is the most commonly used because the satellite series provide nearly 45-year data records with wide-swath coverage, free availability, and relatively medium spatial resolution.[9] However, confusion among various urban land features based on factors such as the spectral and spatial resolution of the data and technique employed has made it impossible to yield desirable results.[10] Weber[11] earlier pointed out that due to the complexity of the urban landscape, classification from satellite images is a difficult task because these images do not exhibit a unique and distinguishable spectral response. On the other hand, high spatial resolution images [e.g., IKONOS, Quickbird, and WorldView (WV)-2/3] have provided new research opportunities especially with the growing demands for monitoring most exceptional urban objects such as building or rooftop footprints and road which were difficult to identify in medium resolution images such as Landsat.[12,13] The primary challenge with using high-resolution images is cost, but when available, it provides useful information for urban assessment and monitoring as compared to moderate resolution satellite images.

According to Sun et al.,[14] the numerous ways to estimate the extent and quantities of built-up areas from medium resolution imagery (e.g., Landsat) in the previous studies can be grouped into five categories: pixel/object-based classification, spectral mixture analysis (SMA), regression model, decision tree, and spectral index-based segmentation. They further asserted that though these classification methods have been widely used, there are challenges and limitations for applying at regional and global scales, e.g., subjective scene-to-scene data analysis, time-consuming, and complicated computing.[7] The limitation of pixel-based approach which is commonly used but does not account for the spatial pattern (i.e., image texture, pixel proximity, feature size, and shape) and mixed-pixel problems in the classification.[15] The object-oriented approach treats an image as a set of significant objects, which requires spatial, spectral, and texture characteristics.[16,17] These requirements create difficulty in optimizing segmentation parameters that hinders the application of object-oriented classification to a large area.[9] Although, SMA-based methods have proven useful for handling the mixed-pixel problems in medium resolution imagery, built-up or impervious surface area (ISA), are commonly overestimated in areas with low-density urban features and underestimated in high-density urban areas.[15,18] SMA-based methods are also not suitable for large-area mapping due to difficulties in end-members related to inter-class variability quantification, and complicated implementation process.[19–21] Also, the regression models (e.g., decision tree) limitations are associated with model calibration and validation from medium resolution images and high-quality ISA reference data and extrapolation of the models in other study areas.[18] Although decision tree is a rule-based method that can effectively process large, high dimensional and nonlinear data, which is suitable for large-area built-up mapping, it is more sensitive to noise and depends significantly on the quality of sample data.[15] Thus, all of the above mentioned are semi-automated and involves human intervention such as a manual selection of representative samples required by the classifiers and incorporates urban morphology information.[22,23]

Thematic-oriented spectral indices have proven to be of good potential in large-area built-up automated mapping due to their easy implementation, parameter-free, and convenience in practical applications.[20,24] Most of the built-up indices for rapid mapping of built-up areas have been developed for medium resolution imagery.[25] Table 1 lists built-up indices developed from medium resolution multispectral Landsat with their merits and demerits. Sameen and Pradhan[35] reported that most of the studies only developed spectral indices for the rapid extraction of built-up areas from medium resolution satellite images (e.g., Landsat).

Table 1 List of built-up indices.

Formula	Imagery	Merits and demerits	References
$NDBI = [(SWIR1 - NIR)/(SWIR1 + NIR)]$	Landsat	Ineffective in differentiating built-up area from bare land areas and requires modification to enhance its accuracy.	26
$NB1 = (RED * SWIR1)/NIR$	Landsat	Higher positive values of NBI indicated a greater possibility of bare ground areas.	27
$IBI = [NDBI - (NDVI + MNDWI)/2]/(NDBI + (NDVI + MNDWI)$	Landsat	Detect asphalt and concrete surfaces and eliminate other land use or cover classes using the vegetation and water index values.	28
$\text{Improved NDBI} = NDBI - NDVI$	Landsat	High subtracted values between the continuous images of NDBI and NDVI coupled with thresholding technique indicates more probability of having built-up regions	29
$NBI = (SWIR1 * RED)/NIR$	Landsat	Red band spectral response was used to enhance the extraction of the built-up region	30
$NDISI = (T_b - (MNDWI + NIR + SWIR1)/3)/T_b + (MNDWI + NIR + SWIR1)/3)$	Landsat	It can represent the real percentage of the impervious surface but sensitive to seasonality.	
$EBBI = (SWIR1 - NIR)/10 * \backslash \text{radic}(SWIR1 + TIR)$	Landsat	It is terrific in distinguishing between built-up and bare land due to the use of TIR band. It detects mixed regions as built-up in bare land areas in regions with a highly heterogeneous landscape, which exhibit high values in NIR and SWIR because of the drier vegetation.	31
$BAEI = (RED + L)/(GREEN + SWIR1)$	Landsat	Excellent in detecting built-up regions, it also has some portion of confusion between built-up and rock due to similar spectral response. Masking of vegetation was required because of its enhancement with built-up regions due to the red band.	13
$BUI = (2 * ((RED * SWIR2) - (SWIR1 * SWIR1))/((RED + SWIR1) + (SWIR1 + SWIR2))$	Landsat	It reveals high positive values for water, high negative values for bare soil and vegetation, and minor negative values for the built-up area. The main focus is on the last finding, i.e., in values ranging around 0.0. Although, it does not follow the logical boundary of the built-up area values adopted in previously developed indices but allows for the best distinction of constructions.	25
$NBUI = EBBI - (SAVI + ((GREEN - SWIR1)/(GREEN + SWIR1))$	Landsat	It highlights the contrast reflection range and absorption in built-up and bare land areas	32
$MNDISI = (T_S - (MNDWI + NIR + SWIR1)/3)/T_S + (MNDWI + NIR + SWIR1)/3)$	Landsat	It suggests that built-up indices are sensitive to image seasonality, and summer is the best time phase for ISA mapping	14

Table 1 (*Continued*).

Formula	Imagery	Merits and demerits	References
$BSTBI = (W_1BLUE + W_2NIR - GREEN) * SWIR2$ Where, $W_1 = (\lambda GREEN - \lambda BLUE)/(\lambda NIR - \lambda BLUE)$; $W_2 = (\lambda NIR - \lambda GREEN)/(\lambda NIR - \lambda BLUE)$	Landsat	It cannot represent the totality of the industrial zone, i.e., underestimation. Age and covering of dust could also suppress the performance of the index.	33
$ENDISI = BLUE - \alpha * [(SWIR1/SWIR2) + (MNDWI)^2]/$ $BLUE + \alpha * [(SWIR1/SWIR2) + (MNDWI)^2]$ $\alpha = 2 * (BLUE)_{Mean}/[(SWIR1/SWIR2)_{Mean} + ((MNDWI)^2)_{Mean}]$ $MNDWI = (Green - SWIR1)/(Green + SWIR1)$	Landsat	It reduces the impacts of arid land, bare rock, and bare soil on IS extraction effectively (i.e., higher separability degree). It presented high values in some grassland areas, i.e., making impervious surface pixels covered with pervious surface pixels	34

L represents an arithmetic constant equal to 0.3, NDBI = Normalized difference built-up index name, NBI = New Built-up Index, IBI = Index-based Built-Up Index, NBI = New Built-up Index, EBBI = Enhanced Built-Up and Bareness Index, BAEI = Built-up area extraction index, BUI = Built-up index, NBUI = New Built-up Index, Ts = retrieved surface temperature, T_b = brightness temperature of the thermal band, MNDWI = modified normalized difference water index, NDISI = normalized difference impervious surface index, MNDISI = Modified Normalized Difference Impervious Surface Index. SAVI = Soil-adjusted Vegetation Index, BSTBI = blue steel tile-roofed buildings index, ENDISI = Enhanced Normalized Difference Impervious Surfaces index, $(.)_{Mean}$ = image mean value, red, blue, green, NIR, SWIR1, and SWIR2 = wavelength of surface reflectance bands.

Therefore, with the growing demands to observe, assess, and monitor land change processes and biophysical characteristics of urban environments more accurately, it is necessary to develop new built-up indices for recently available multispectral high spatial resolution satellite imagery such as IKONOS, Quickbird, and WV-2/3.[36] Based on the problems mentioned above in index-based built-up (IBI) area mapping, there is a need to develop new indices that are accurate and able to accommodate the spectral variability of the built-up materials. The objective of the study is to propose and test the utility of a new built-up extraction index (NBEI) based on WV-2 imagery. The key research questions are:

1. Which WV-2 spectral bands are essential in characterizing built-up areas, i.e., consider the spectral variability of the built-up materials (e,g., types, age, and colour)?
2. Could NBEI threshold analysis eliminate non built-up interference to ensure high accuracy for build-up ISA extraction?
3. In a comparative analysis with previous spectral indices, will NBEI perform better?

2 Material and Methods

2.1 *Study Area*

Geographically, the study site (Lynwood Ridge) is located between S25°45′29.66″ E28°17′33.66″ (top left), S25°45′39.29″ E28°17′44.52″ (bottom right) and S25°45′29.57″ E28°17′44.42″ (top right), S25°45′39.39″ E28°17′33.76″ (bottom left) Pretoria, Gauteng Province, South Africa (Fig. 1). It is situated in the eastern part of Pretoria, which is in the north-northeast of Johannesburg, Gauteng Province, South Africa (Fig. 1). It is a suburb known as the old east and boasts having some of the city's most high-end residential properties. Lynwood Ridge has a topography 1336-m above sea level with a total area of 2.13 km^2 situated in the Pretoria city that has an average temperature ranging from 37°C max to 18°C min according to the South African Weather Service, SAWS, 2011. Based on the STATISTICS SA, South African National census, 2011, Lynnwood Ridge has an estimated population of 3463 (1623.15 per km^2). For this comparative study (Fig. 2), Lynnwood Ridge a formal residential area situated in Pretoria was selected.

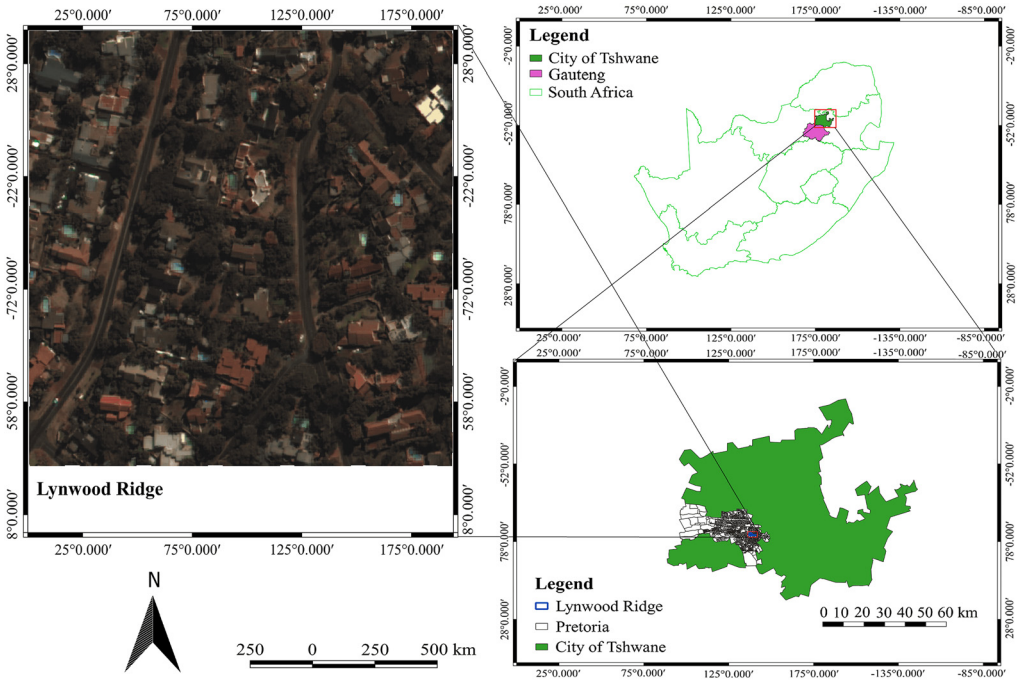

Fig. 1 The location of the study area relative to South Africa.

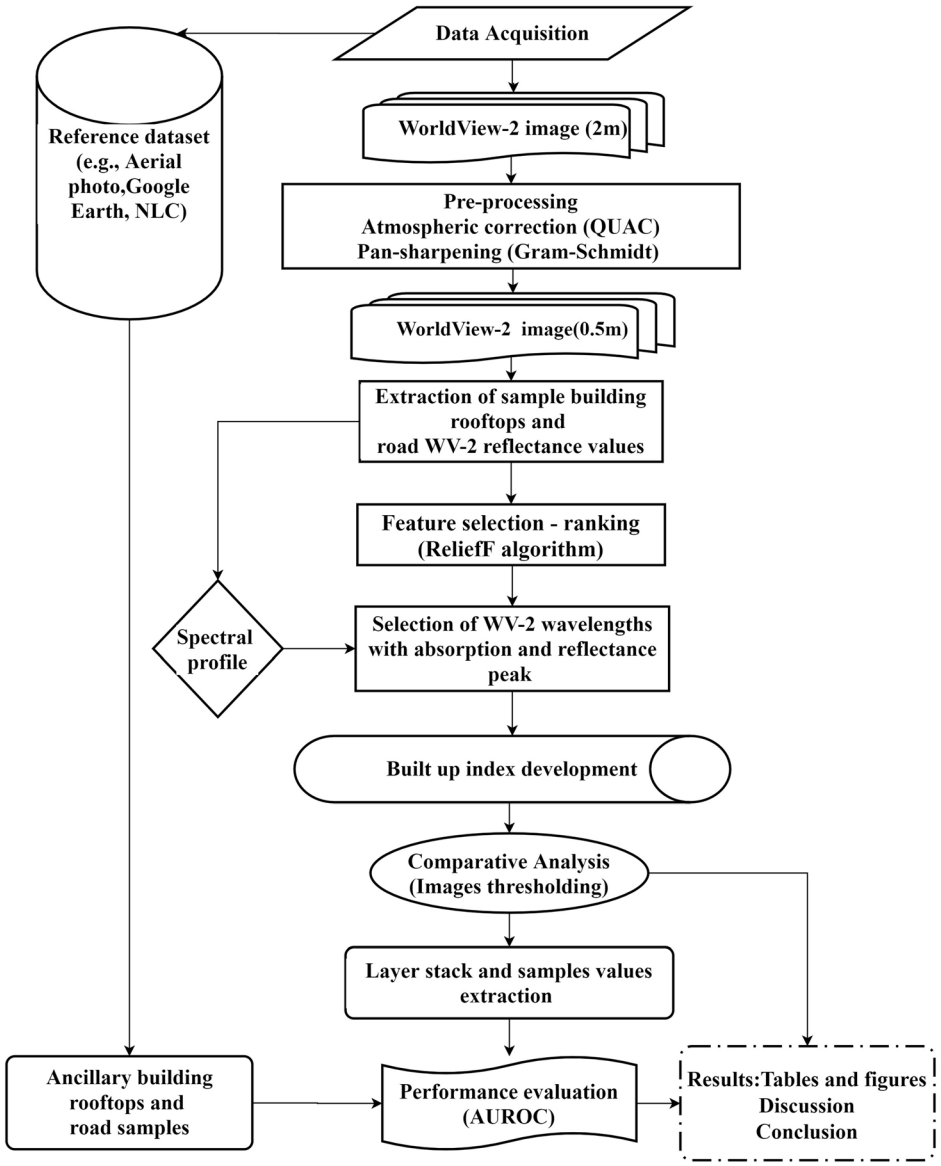

Fig. 2 A flowchart showing the process followed for the formation of the NBEI for WV-2.

2.2 *Study Data*

One cloud-free WV-2 multispectral image of the study area was obtained from Digital Globe in October 2015. WV-2 was launched in October 2009 as the first commercial multispectral satellite that comprises 2-m eight multispectral bands and 0.5-m panchromatic bands.[37] The satellite has a swath width of 16.4 km, an average revisits time of 1.1 days and is capable of collecting up to 9.75×10^5 km^2 per day. Therefore, the satellite has the spectral and spatial resolutions that meet a variety of remote sensing applications such as mapping rooftops.[38,39] The Department of Environmental Affairs, South Africa subcontracted National Land Cover classification dataset of 2013–2014 completed by a consulting company GeoterraImage (Pty), 1:10,000 aerial photography at 0.5-m spatial resolution (November 2015) from CDNGI together with Google Earth search engine were used for validation purpose to verify the rooftop colours in the study site. Software to be used for desktop analysis were; ESRI (ArcGIS), EXELIS Environment for Visualizing Images (ENVI) IDL version 5.2, QGIS, Microsoft Office, WEKA, R statistical software environment (RStudio, Inc., Boston, Massachusetts, Version 1.1.463) software.

Table 2 WV-2 imagery characteristics.

Spectral bands	Wavelength (nm)	Spatial resolution (m)
Coastal blue	400 to 450	2
Blue	450 to 510	2
Green	510 to 580	2
Yellow	585 to 625	2
Red	630 to 690	2
Red-edge	705 to 745	2
NIR1	770 to 895	2
NIR2	860 to 1040	2
Pan	450 to 800	0.5

2.3 *Image Acquisition and Pre-Processing*

The WV-2 scene used in this study was orthorectified and geometrically corrected by Digital Globe.[40–42] The WV-2 bands (Table 2) were then atmospherically corrected and transformed into canopy reflectance using the Quick Atmospheric Correction extension in Environment for Visualizing Images (ENVI) software (ENVI, 2015). After the radiometric calibration, the WV-2 multispectral bands pan-sharpening was performed using the ENVI Gram–Schmidt Spectral Sharpening algorithm that improved the visualization of rooftop material.[43] Finally, the WV-2 image was then referenced to the Universal Transverse Mercator (UTM zone 35 South) projection using WGS-84 Geodetic datum.[44]

3 Feature Selection Using ReliefF Algorithm

3.1 *Feature Selection*

Feature selection is an important pre-processing step in pattern recognition and machine learning, artificial intelligence and data mining communities. It helps us to focus the attention of a classification algorithm on those features or bands that are the most relevant to predict the classes (i.e., built-up).[45] Based on statistical distribution using a large number of features as the inputs of initiating algorithms have the disadvantage (e.g., inefficient as it consumes memory and time, irrelevant features may confuse classification algorithms) and advantages (e.g., improving understandability and lowering the cost of data acquisition and handling). Feature selection methods can be grouped into two categories, which are ranking features according to the same evaluation criterion and choosing a minimum set of features that satisfies an evaluation criterion. In this work, we implemented the ReliefF algorithm for feature ranking in the WEKA software to identify the significant wavelengths sensitive to the built-up areas (e.g., road and rooftops).

3.2 *Algorithm Implementation*

In this study, we employed the use of the robust ReliefF algorithm because of its ability to handle both a binary and multiclass problems and accommodate an incomplete and noisy data.[46] Based on the generated training or stratified random sample points ($n = 130$) on built-up areas (e.g., roads and rooftops), we first extracted from the eight WV-2 bands (i.e., features) reflectance values using the "extraction" module of the spatial analyst tool of ArcGIS software. Before this step, the validity of all these sample points selected was carefully checked through visual inspection of both aerial photography (November 2015) as well as Google Earth engine to avoid mislabeling problems. Subsequently, the ReliefF algorithm was implemented in the WEKA software

using the k-nearest neighbors (KNN). The ReliefF algorithm randomly selects from the training a sample n_i, and then performs searches for KNN in two ways, i.e., from the same class (nearest hits h_j) and a different class, called nearest misses $m_j(C)$ [see Eqs. (1) and (2)]. It updates the quality estimation $W[A]$ for all attributes "A" depending on the WV-2 reflectance values associated with the training sample (n_i) hits and misses. The contribution for each class of the misses is weighted with the prior probability of that class $p(C)$ (estimated from the sample points). Since we want the contributions of hits and misses in each step to be in 0 to 1 symmetric, we ensured that misses probability weights sum to 1. As the class of hits is missing in the sum, we divided each probability weight with factor $1 - p$ (class (n_i)) (which represents the sum of probabilities for the misses). The process is repeated for t times. Selection of k hits misses and ensured greater robustness of the ReliefF algorithm about the noise. k parameter that controls the locality of the estimates was set to 10.[45,46] Finally, the ReliefF algorithm ranked the WV-2 bands sensitiveness to the built-up areas based on the estimated weight ranging from 0 to 1 with large weights assigned to important WV-2 bands. ReliefF algorithm was used successfully for feature extraction.[45,47,48]

3.3 *Algorithm Representation*

The input: for each training instance a vector or point of the attribute (i.e., either road or rooftop)

Algorithm 1

Output: the vector w of estimations of the qualities of attributes.

1. set all weights $w[A] := 0.0$;
2. for $i := 1$ to t do begin
3. randomly select an instance n_i;
4. find k-nearest hits h_j;
5. for each class $C \neq$ class (n_i) do
6. from class C find k nearest misses m_j (c);
7. for $A := 1$ to a do
8. $$w[A] = w[A] - \sum_{j=1}^{k} \frac{diff(\mathrm{A}, \mathrm{ni}, \mathrm{hj})}{(m, k)} + \sum_{C \neq classn_i} \frac{\left[\frac{p(\mathrm{c})}{1-p(\mathrm{class(ni)})} \sum_{j=1}^{k} diff(\mathrm{a}, \mathrm{nihj})\right]}{(\mathrm{m}, \mathrm{k})} \quad (1)$$
9: End

4 Built-Up Spectral Index Creation

4.1 *Spectral Index Creation*

Spectral indices are part of processing methods called multi-spectral transformations.[13] Caloz et al.[49] earlier defined an index as a variable synthetic, digital characterizing the intensity or the extension of an overly complex phenomenon to be broken down into a manageable number of parameters. The stratified random sampling technique was used to develop an empirical method to formulate the built-up index. Deng et al.[7] asserted that the selection of stratified random samples is vital for the construction of the spectral index. Gao and Mas[50] pointed out that in the selection of samples, the large number represents the reliability of specific land cover spectral signature much better than small numbers. Therefore, in this study as a compromise of efficiency

and reliability, the generated stratified random sample points or pixels on built-up materials used in the above feature ranking were comprehensively used to examine their spectral pattern. In the development of the NBEI we employed the following steps:

1. Evaluating the spectral pattern to identify significant WV-2 wavelengths that are sensitive to dissimilarities in a target's spectral response, which represent the built-up areas as compared to other land features.[51]
2. We postulate that the significant WV-2 bands obtained from the feature ranking using the ReliefF algorithm will be the same spectral bands that show the absorption and reflectance regions of the spectral pattern.
3. Since built-up samples may exhibit a high spectral variability due to their complex materials, we hypothesize that an effective spectral index will be developed based on the ratios (i.e., band combinations) between the sum of spectral bands that signifies the absorption and reflectance regions of the spectral pattern.

4.2 *Comparative Analysis*

To better examine the performance of NBEI in creating a distinction between built-up and non-built-up features in WV-2 imagery, we also conducted a comparative analysis. In this study, five related spectral indices: built-up area index (BAI),[52] built-up the spectral index (BSI),[35] red edge/green index (RGI),[53] and WorldView built-up index (WV-BI)[54] were implemented for the comparative analysis since they could enhance built-up features information with a degree of vagueness. The empirical formulas of these indices are expressed in Table 3 below. Finally, the threshold technique employed in the study for the separation of built-up and non-built-up areas, i.e., to determine the optimal threshold value follows the same steps employed in the study of Refs. 28 and 30.

4.3 *Precision Evaluation*

In this study, the accuracy of the five BSIs was assessed based on un-stratified random evenly distributed samples ($n = 244$) obtained from reference data (aerial photo at 0.5-m spatial resolution and visual inspection on Google Earth) to validate the index derived image. 10-fold cross-validation was used to rearrange the samples to ensure that each fold is a good representation of the whole datasets, i.e., with a lower sample distribution variance compared to the hold-out cross-validation.[55] Finally, we implemented the performance evaluation metrics using the area under the receiver operating characteristic curve (AUROC) which is a graph that summarizes the performance of the indices (classifier) over all possible thresholds. It is generated by plotting the true positive rate (y axis) against the false positive rate (x axis). Wieland and Pittore[47] further explained that the true positive rate is the proportion of actual positives that are classified as positives, while true negative rate, is the proportion of actual negatives, which are classified as negatives. It was computed using InformationValue and plotROC package in the R statistical software environment.[35,56,57]

Table 3 List of compared WV-2 images built-up indices.

Name	Formula	Reference
BAI	[(Blue – NIR1)]/ [(Blue + NIR1)]	52
BSI	[(Yellow – 2*NIR1)/ (Yellow + 2*NIR1)]	35
RGI	[(red edge – green)/ (red edge + green)]	53
WV-BI	[(coastal blue – red edge)/ (coastal blue + red edge)]	37, 54
NBEI	[(NIR2 +NIR1) – (green + red edge)]/ [(NIR2 +NIR1) + (green + red edge)]	This study

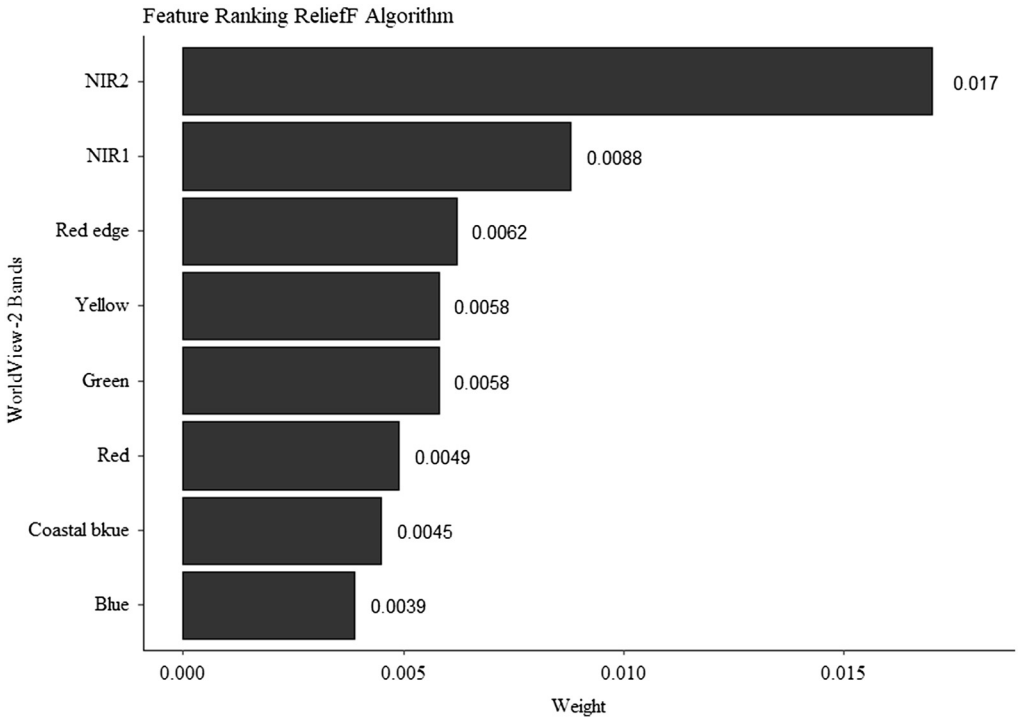

Fig. 3 WV-2 spectral bands features ranked by their importance for identifying built-up areas.

5 Results

5.1 *An Empirical Analysis of ReliefF Algorithm*

Figure 3 shows an overview of the WV-2 bands importance for delineating built-up using the ReliefF algorithm. As shown in Fig. 3, NIR2 is ranked the highest with weight of 0.017. Amongst the eight spectral band features, the most important features that can be used for the formulation of the new normalized band ratio for built-up area delineation were NIR2, NIR1, red edge, and green or yellow.

5.2 *Reflectance Profile of Built-Up Surfaces and Spectral Index Creation*

The concept for developing spectral indices is to identify the weakest and strongest reflectance band from multi-spectral data.[58] Figure 4 shows that the built-up (i.e., rooftop and roads) reflectance is distinct. As shown in Fig. 4, the spectral profiles of rooftop and road exhibit a similar reflectance pattern of shape but differ largely in the magnitude particularly in the red edge to NIR spectrum. The built-up surfaces indicated reflection highly in the NIR (centered between 833 and 950 nm) regions and absorption at the green (545 nm) and red edge (725 nm) regions of the WV-2.

5.3 *Spectral Index Creation and Threshold Selection for Built-Up Mapping*

Based on the identified weakest and strongest bands coupled with conventional approaches, Waqar et al.[10] explained that spectral indices are formed to enhance required land cover over wide range of wavelength values and suppress others. To develop the NBEI, the WV-2 bands NIR2, NIR1, red edge, and green were used, respectively. NBEI was developed using normalized ratio of the addition of NIRs and green and red edge spectral regions which is mathematically expressed as

$$NBEI = \frac{((NIR2 + NIR1) - (Green + Red\,edge))}{((NIR2 + NIR1) + (Green + Red\,edge))}. \quad (2)$$

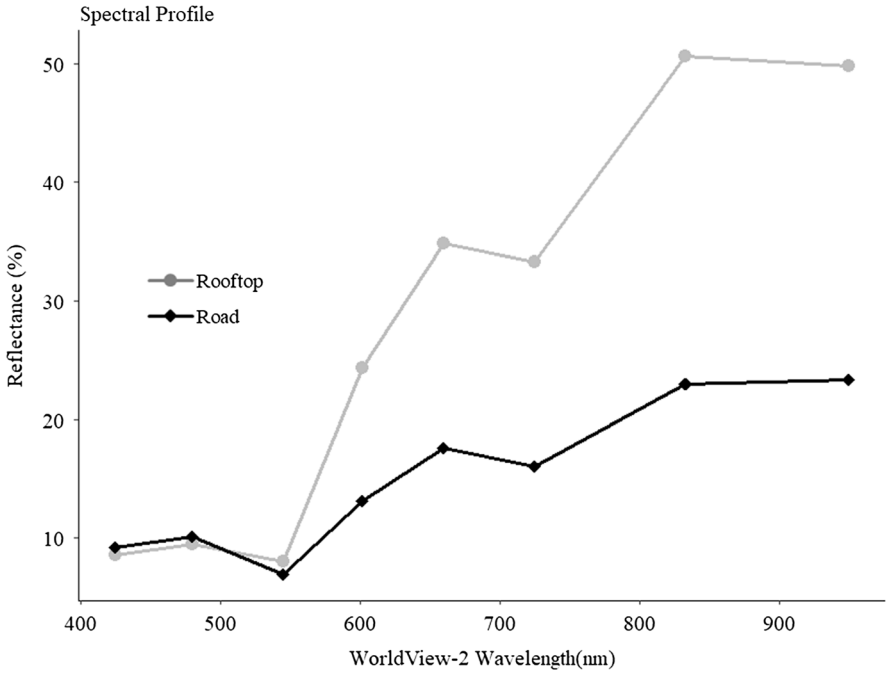

Fig. 4 WV-2 mean spectral profiles of built-up areas extracted.

The NBEI can enhance the built-up land feature easily because the subtraction of the green and red edge bands from the NIR bands will result in positive values for built-up land pixels only, i.e., the index takes advantage of the condition where the features with higher NIR2 and NIR1 values but lower Green and Red edge values. Likewise, the subtraction of the green and red edge bands from the NIR bands in the band combination helped enhance the depiction of built-up area while suppressing other non-built-up areas.[59] Evidently, the NBEI is a normalized difference index with features such as: (1) a ratio-based index, (2) values ranging between -1 and $+1$, and (3) enhanced built-up information has positive values, while the suppressed land covers (e.g., vegetation, base land, or soil) has zero to negative values.

Furthermore, based on threshold analysis of binary images (Fig. 6) derived from previously developed built-up indices, the results of the comparative study in Fig. 5 indicate the following threshold values in Table 4, respectively.

In comparison, our developed index (i.e., NBEI) demonstrated its potential to map built-up rooftops and asphaltic roads based on the produced index threshold values in Table 4 and Fig. 5. The stratified 10-fold cross-validation indicate the AUROC of NBEI improved the detection of built-up (AUROC $= \sim 0.82$) as compared to the existing ones such as BSI (AUROC $= \sim 0.78$), BAI (AUROC $= \sim 0.73$), RGI (AUROC $= \sim 0.71$), and WVBI (AUROC $= \sim 0.67$), respectively (Fig. 8). The threshold values indicate built-up or impervious surfaces (i.e., $n = 130$ stratified random samples of built-up pixels). Threshold values within the range of 0.03 to 0.509 (see Table 4), are primarily clustered values with histogram frequencies of 5 to 20. Threshold values within ranges of <0.03 and >0.509 are mostly frequency that optimizes non-built-up areas. Overall, our results shows that NBEI has successfully separated built-up areas from other land cover types with relatively high precision.

6 Discussion

According to Varshney and Rajesh,[8] index-based algorithms it is possible to classify built-up areas such as roads and rooftops and so on. Automatically at the minimal time when compared to the conventional image classification process. Due to the convenience of spectral indices in detecting specific land cover, a large number of indices have been developed in the past decade.[7] However, few spectral indices are available to enhance built information directly from the very

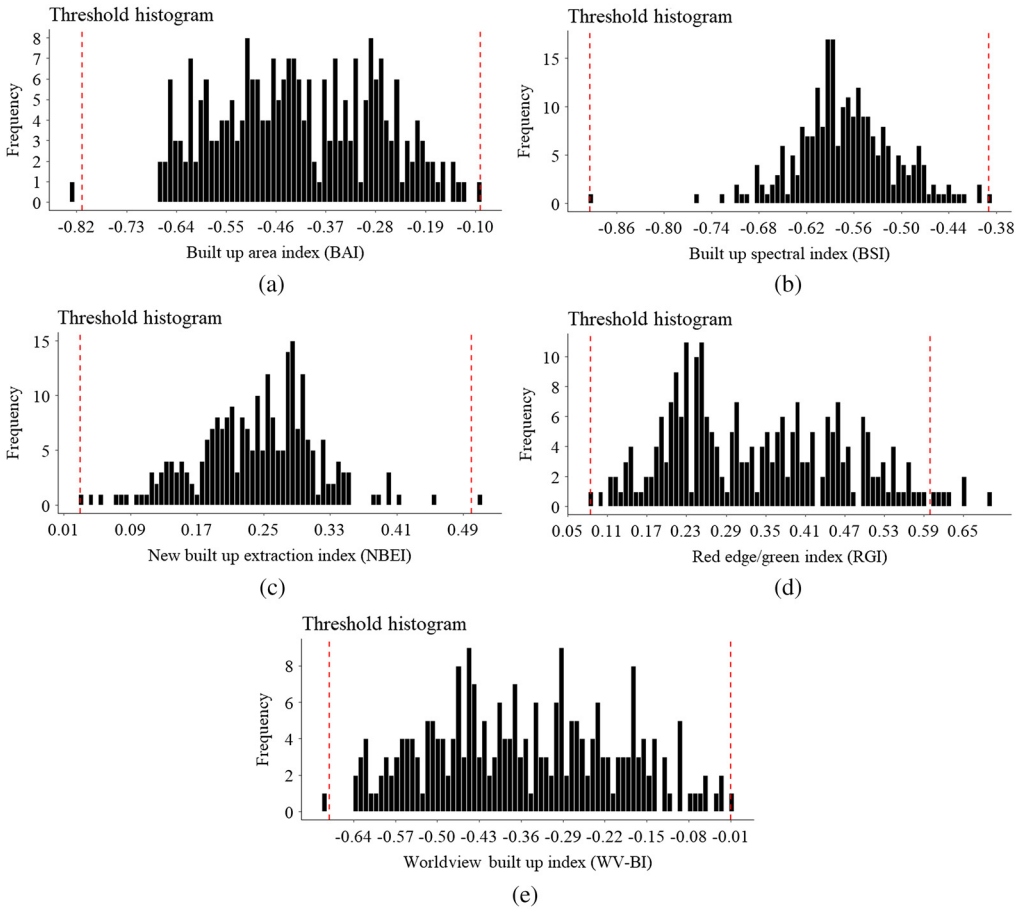

Fig. 5 Histogram showing the range of the threshold pixel values depicting built-up areas in (a) BAI; (b) BSI; (c) proposed (NBEI); (d) RGI; (e) WV-BI.

high resolution image (e.g., WV-2). This study proposed NBEI for built-up extraction based on WV-2 bands. Accuracy assessment results show that overall extraction accuracy for built-up areas using NBEI is greater than that of previously developed indices. This shows that our proposed NBEI can be used to achieve much better results for extraction and better delineating of built-up areas in our study area.

In this study, we examined the capability of NBEI to identify different roads and to build rooftops materials. Amongst the eight WV-2 spectral bands, the significant bands used for the formulation of the new index for built-up area detection were NIR2, NIR1, red edge, and green. These significant wavelengths were able to depict a significant difference in the shape of spectral signature.[60] The NIR1 and NIR2 bands located at the reflectance region and the Green band at the absorption region helped to enhance built-up areas[51] while the red edge band suppresses the vegetation.[61,62] Although previous WV-2 built-up indices employed for comparison were based on various bands (i.e., coastal blue, blue, green, yellow, red edge, and NIR1), the ReliefF algorithm employed in our approach, found that green, red edge, NIR1, and NIR2 are the most suitable bands to extract the built-up areas from the WV-2 image. Also, the visual examination of these spectral indices performance presented in Fig. 7, further explains the superior performance obtained of NBEI. For instance, the low accuracy of BAI, RGI, and WV-BI (AUROC ≤ 0.73) could be attributed to the fact that many rooftops (e.g., red and brown) were not detected accurately. This might be because of the absence of NIR2 in these indices that can comprehend the different rooftops surface illumination. Also, the subtraction of the green and red edge bands from the NIR bands in the band combination was helpful in enhancing the depiction of built-up area while suppressing other non-built-up areas,[59] i.e., vegetation and bare land or soil. BSI performed reasonably well with negative threshold values when compared to NBEI which had good coverage of built-up areas with positive threshold values. Even though the study

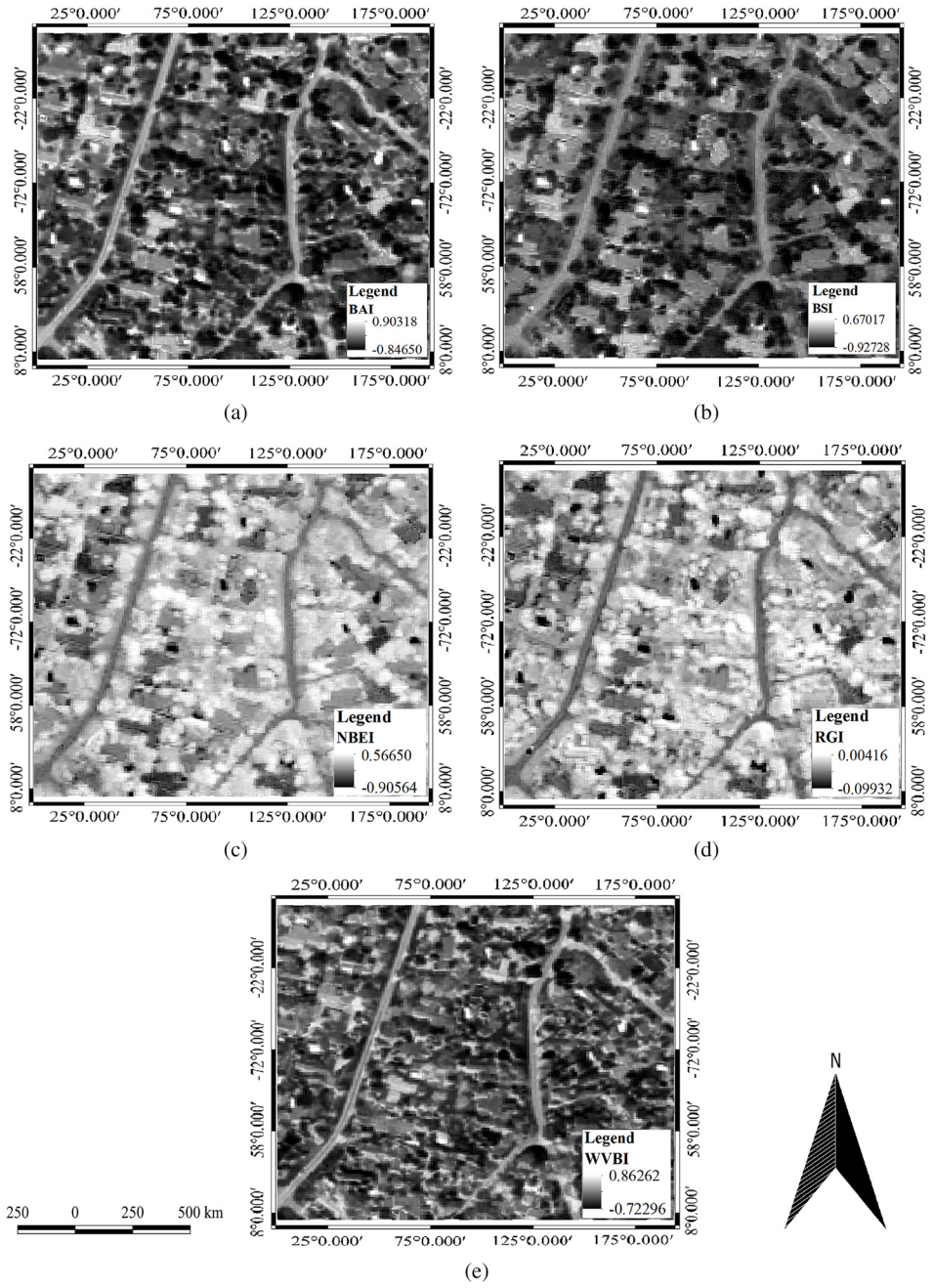

Fig. 6 Result of compared spectral indices images (a) BAI; (b) BSI; (c) proposed (NBEI); (d) RGI; (e) WV-BI.

Table 4 Dynamic ranges of previous WV-2 built-up indices.

Index	Threshold
BAI	−0.09 ≥ built-up ≥ −0.81
BSI	−0.390 ≥ built-up ≥ −0.894
NBEI	0.03 ≥ built-up ≥ 0.509
RGI	0.085 ≥ built-up ≥ 0.69
WV-BI	−0.009 ≥ built-up ≥ −0.689

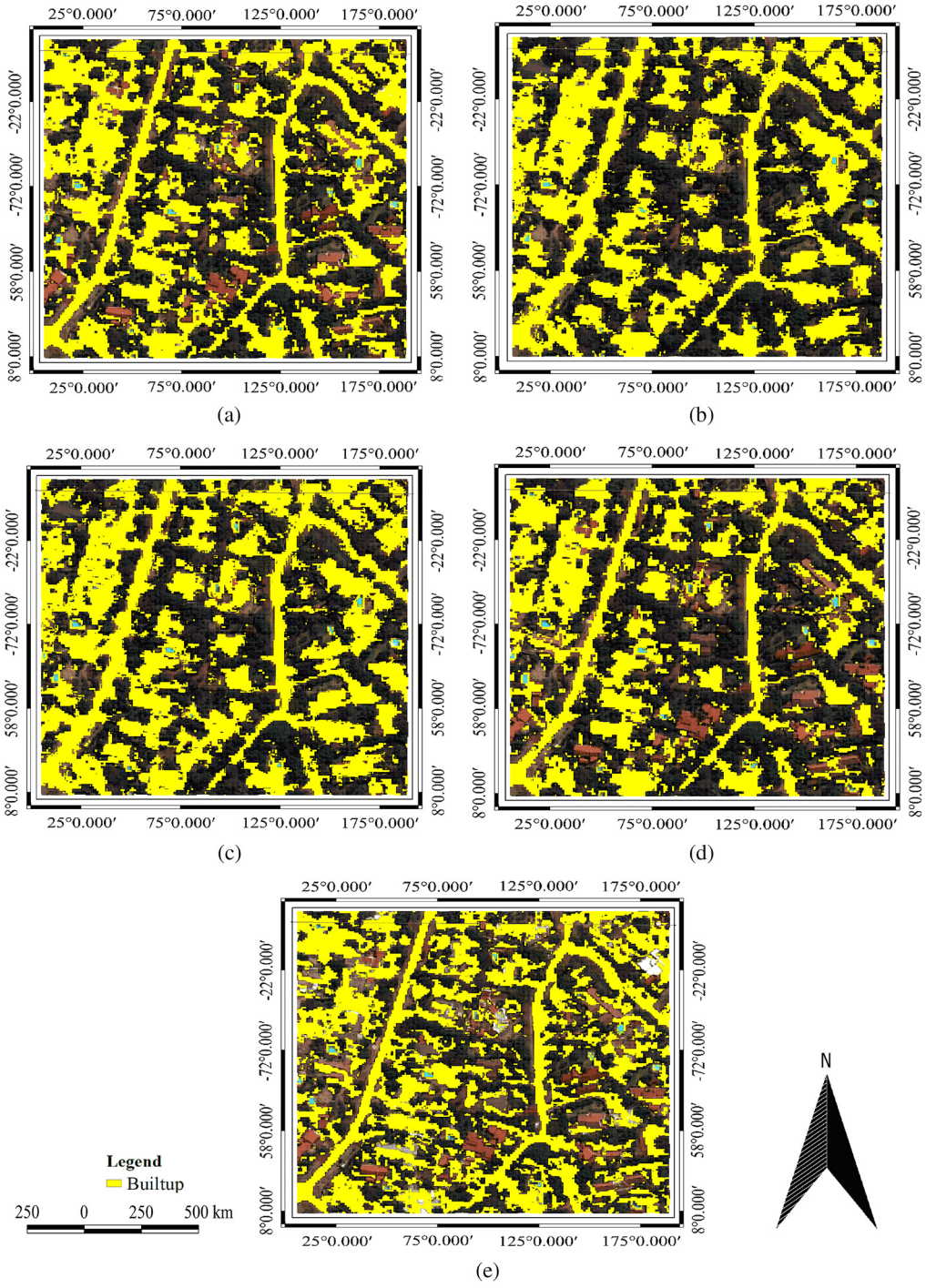

Fig. 7 Thematic output after thresholding of (a) BAI; (b) BSI; (c) proposed (NBEI); (d) RGI; (e) WV-BI.

area comprises a heterogeneous natural (e.g., vegetation, bare land, or soil) and artificial land cover, NBEI developed in this study provides the advantage of independence from making a mask for non-built-up to focus only on built-up areas.

7 Conclusion

The new spectral index (NBEI) proposed in this study improves the extraction of built-up areas automatically from WV-2 imagery. The evaluation of the NBEI compared with four other

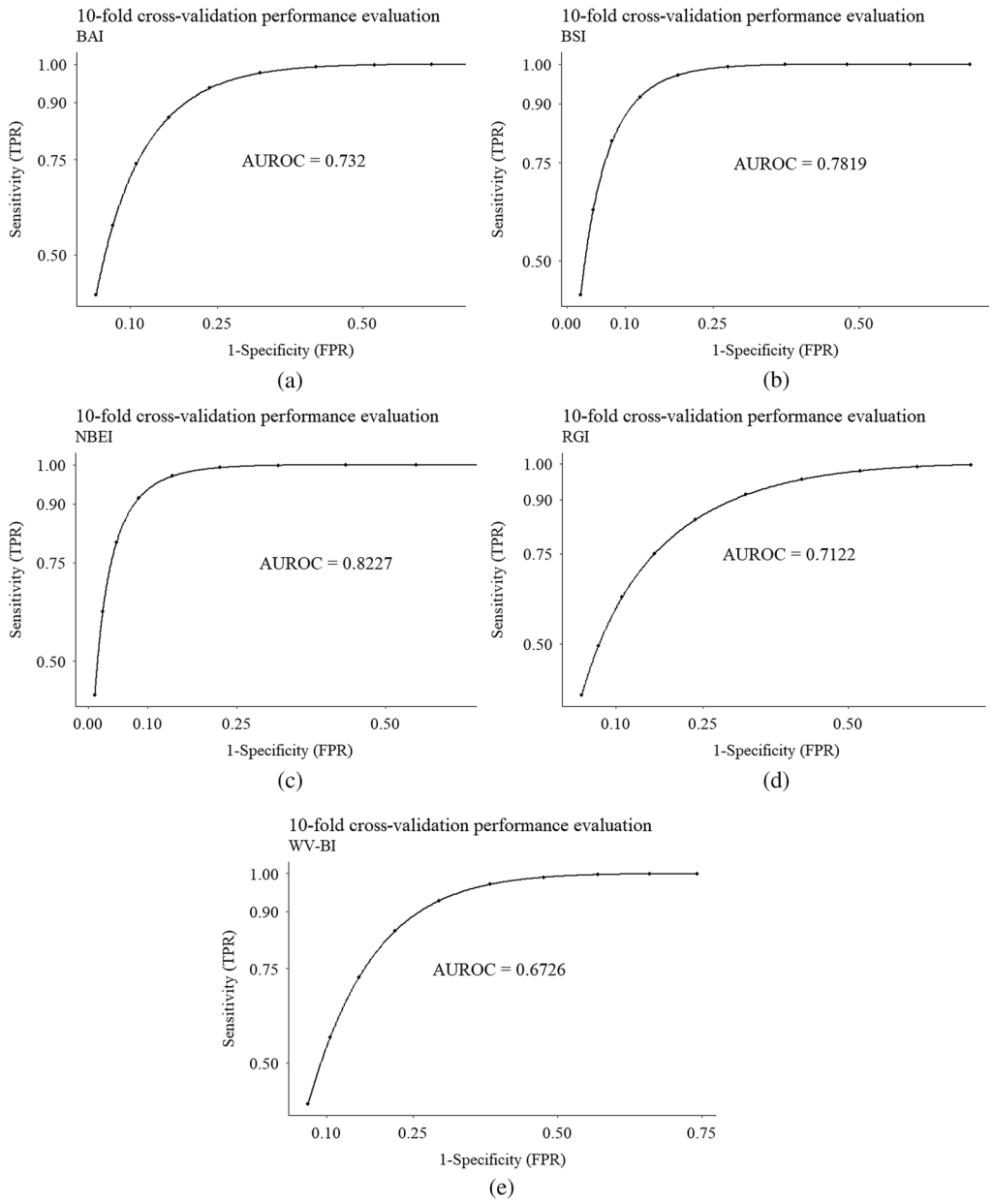

Fig. 8 AUROC curve showing the performance of all the built-up indices of WV-2; (a) BAI; (b) BSI; (c) proposed NBEI; (e) RGI; (d) WV-BI.

spectral built-up indices developed for WV-2 imagery in the previous study showed a better result for the detection and extraction of built-up areas with less time and pre-processing. Since the performance of the spectral index depends on the spectral response of land cover characteristics that vary from one region to another due to climatic, topographic and socio-economic changes, the effectiveness of the NBEI for WV-2 still needs to be tested at different study sites within the urban settings (e.g., commercial and industrial), provided there is the availability of this commercial satellite imagery.

Acknowledgments

The authors would like to thank the University of South Africa Student Funding Directorate (UNISA, DSF) for providing the research bursary to conduct this study. Conflicts of Interest: The authors declare no conflict of interest.

References

1. H. Xu, "Analysis of impervious surface and its impact on Urban heat environment using the normalized difference impervious surface index (NDISI).," *Photogramm. Eng. Remote Sens.* **76**(5), 557–565 (2010).
2. UNDP, *World Urbanization Prospects: The 2011 Revision*, UNDP, New York (2012).
3. J. Odindi, P. Mhangara, and V. Kakembo, "Remote sensing land-cover change in Port Elizabeth during South Africa's democratic transition," *S. Afr. J. Sci.* **108**(5–6), 60–66 (2012).
4. M. R. Montgomery and P. C. Hewett, "Urban poverty and health in developing countries: household and neighborhood effects," *Demography* **42**(3), 397–425 (2005).
5. A. Adeyemi et al., "Effect of impervious surface area and vegetation changes on mean surface temperature over Tshwane metropolis, Gauteng Province, South Africa," *South African J. Geomatics* **4**(4) (2015).
6. S. T. Jennings and D. Jarnagin, "Impervious surfaces and stream flow discharge: a historical remote sensing perspective in a northern Virginia sub watershed," in *ASPRS Annu. Conf.*, Washington, D.C. (2000).
7. Y. Deng et al., "RNDSI: a ratio normalized difference soil index for remote sensing of urban/suburban environments," *Int. J. Appl. Earth Obs. Geoinf.* **39**, 40–48 (2015).
8. A. Varshney and E. Rajesh, "A comparative study of built-up index approaches for automated extraction of built-up regions from remote sensing data," *J. Indian Soc. Remote Sens.* **42**(3), 659–663 (2014).
9. S. Yu et al., "Monitoring and analyzing the spatial dynamics and patterns of megacities along the Maritime Silk Road," *Yaogan Xuebao/J. Remote Sens.* **21**(2), 169–181 (2017).
10. M. M. Waqar et al., "Development of new indices for extraction of built-up area and bare soil from landsat data," *J. Geophys. Remote Sens.* **1**(1), 1–4 (2012).
11. C. Weber, Remote Sensing Data Used for Urban Agglomeration Delimitation, in *Remote Sensing and Urban Analysis*, Taylor & Francis, London (2001).
12. M. Bouziani, K. Goïta, and D. C. He, "Change detection of buildings in urban environment from high spatial resolution satellite images using existing cartographic data and prior knowledge," in *Int. Geosci. Remote Sens. Symp.*, pp. 2581–2584 (2007).
13. S. Bouzekri, A. A. Lasbet, and A. Lachehab, "A new spectral index for extraction of built-up area using Landsat-8 data," *J. Indian Soc. Remote Sens.* **43**(4), 867–873 (2015).
14. Z. Sun et al., "A modified normalized difference impervious surface index (MNDISI) for automatic urban mapping from landsat imagery," *Remote Sens.* **9**(9), 942 (2017).
15. Z. Sun, "Estimating urban impervious surfaces from Landsat-5 TM imagery using multilayer perceptron neural network and support vector machine," *J. Appl. Remote Sens.* **5**(1), 053501 (2011).
16. J. Qian, Q. Zhou, and Q. Hou, "Comparison of pixel-based and object-oriented classification methods for extracting built-up areas in aridzone," in *ISPRS Workshop on Updating Geo-Spatial Databases with Imagery & The 5th ISPRS Workshop on DMGISs*, Vol. 8(August), pp. 163–171 (2007).
17. G. Yan et al., "Comparison of pixel-based and object-oriented image classification approaches—a case study in a coal fire area, Wuda, Inner Mongolia, China," *Int. J. Remote Sens.* **27**(18), 4039–4055 (2006).
18. Q. Weng, "Remote sensing of impervious surfaces in the urban areas: requirements, methods, and trends," *Remote Sens. Environ.* **117**, 34–49 (2012).
19. F. Yang, B. Matsushita, and T. Fukushima, "A pre-screened and normalized multiple endmember spectral mixture analysis for mapping impervious surface area in Lake Kasumigaura Basin, Japan," *ISPRS J. Photogramm. Remote Sens.* **65**(5), 479–490 (2010).
20. C. Deng and C. Wu, "A biophysical composition index for remote sensing of urban environments," *Remote Sens. Environ.* **127**, 247–259 (2012).
21. D. Lu et al., "Methods to extract impervious surface areas from satellite images," *Int. J. Digital Earth* **7**(2), 93–112 (2014).
22. L. T. Waser et al., "Evaluating the potential of worldview-2 data to classify tree species and different levels of ash mortality," *Remote Sens.* **6**(5), 4515–4545 (2014).

23. C. Wei and T. Blaschke, "Pixel-wise vs. object-based impervious surface analysis from remote sensing: correlations with land surface temperature and population density," *Urban Sci.* **2**(1), 2 (2018).
24. G. Sun et al., "Combinational build-up index (CBI) for effective impervious surface mapping in urban areas," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **9**(5), 2081–2092 (2016).
25. D. Kaimaris and P. Patias, "Identification and area measurement of the built-up area with the built-up index (BUI)," *Int. J. Adv. Remote Sens. GIS* **5**(1), 1844–1858 (2016).
26. Y. Zha, S. Ni, and S. Yang, "Use of normalized difference built-up index in automatically mapping urban areas from TM imagery," *Int. J. Remote Sens.* **24**, 583–594 (2003).
27. X. L. Chen et al., "Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes," *Remote Sens. Environ.* **104**(2), 133–146 (2006).
28. H. Xu, "A new index for delineating built-up land features in satellite imagery," *Int. J. Remote Sens.* **29**(14), 4269–4276 (2008).
29. H. Chunyang et al., "Improving the normalized difference built-up index to map urban built-up areas using a semiautomatic segmentation approach," *Remote Sens. Lett.* **1**, 213–221 (2010).
30. J. Chen et al., "Extract residential areas automatically by new built-up index," in *18th Int. Conf. Geoinf.* (2010).
31. A. R. As-Syakur et al., "Enhanced built-UP and bareness index (EBBI) for mapping built-UP and bare land in an urban area," *Remote Sens.* **4**(10), 2957–2970 (2012).
32. P. Sinha, N. K. Verma, and E. Ayele, "Urban built-up area extraction and change detection of adama municipal area using time-series Landsat images," *Int. J. Adv. Remote Sens. GIS* **5**(1), 1886–1895 (2016).
33. Z. Guo et al., "A new index for mapping the 'blue steel tile' roof dominated industrial zone from Landsat imagery," *Remote Sens. Lett.* **9**(6), 578–586 (2018).
34. J. Chen et al., "Enhanced normalized difference index for impervious surface area estimation at the plateau basin scale," *J. Appl. Remote Sens.* **13**(1), 016502 (2019).
35. M. I. Sameen and B. Pradhan, "A novel built-up spectral index developed by using multi-objective particle-swarm-optimization technique," *IOP Conf. Ser. Earth Environ. Sci.* **37**(1), 012006 (2016).
36. S. Myint et al., "Combined effects of impervious surface and vegetation cover on air temperature variations in a rapidly expanding desert city," *GIScience Remote Sens.* **47**(3), 301–320 (2010).
37. A. Wolf, *Using WorldView 2 Vis-NIR MSI Imagery to Support Land Mapping and Feature Extraction Using Normalized Difference Index Ratios*, Longmont, Colorado (2010).
38. N. E. M. Nasarudin and H. Z. M. Shafri, "Development and utilization of urban spectral library for remote sensing of urban environment," *J. Urban Environ. Eng.* **5**(1), 44–56 (2011).
39. E. Taherzadeh and H. Z. M. Shafri, "Development of a generic model for the detection of roof materials based on an object-based approach using WorldView-2 Satellite imagery," *Adv. Remote Sens.* **2**(4), 312–321 (2013).
40. O. Mutanga et al., "Evaluating the robustness of models developed from field spectral data in predicting African grass foliar nitrogen concentration using WorldView-2 image as an independent test dataset," *Int. J. Appl. Earth Obs. Geoinf.* **34**(1), 178–187 (2015).
41. S. Gairola et al., "Remote sensing object-oriented approaches coupled with ecological informatics to map invasive plant species," *South Afr. J. Geomat.* **5**(3), 285 (2016).
42. S. Madonsela et al., "Multi-phenology WorldView-2 imagery improves remote sensing of savannah tree species," *Int. J. Appl. Earth Obs. Geoinf.* **58**, 65–73 (2017).
43. C. Laben and B. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," US Patent 6011875A (2000).
44. G. Omer et al., "Performance of Support vector machines and artificial neural network for mapping endangered tree species using WorldView-2 data in Dukuduku Forest, South Africa," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **8**(10), 4825–4840 (2015).
45. R. P. L. Durgabai and Y. Ravi Bhushan, "Feature selection using relieff algorithm," *IJARCCE* **3**(10), 8215–8218 (2014).

46. M. Robnik and I. Konenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Mach. Learn.* **53**(1-2), 23–69 (2003).
47. M. Wieland and M. Pittore, "Performance evaluation of machine learning algorithms for urban pattern recognition from multi-spectral satellite images," *Remote Sens.* **6**(4), 2912–2939 (2014).
48. K. Anusha and E. Sathiyamoorthy, "Comparative study for feature selection algorithms in intrusion detection system," *Autom. Control Comput. Sci.* **50**(1), 1–9 (2016).
49. R. Caloz and C. Collet, *Précis de télédétection – Volume 3: Traitements numériques d'images de télédétection*, p. 386, Presses Universitaires du Québec-Agence Universitaire de la Francophonie (2001).
50. Y. Gao and J. Mas, "A comparison of the performance of pixel based and object based classifications over images with various spatial resolutions," *Online J. Earth Sci.* **2**(8701), 27–35 (2008).
51. S. H. Samsudin, H. Z. M. Shafri, and A. Hamedianfar, "Development of spectral indices for roofing material condition status detection using field spectroscopy and WorldView-3 data," *J. Appl. Remote Sens.* **10**(2), 025021 (2016).
52. P. Mhangara et al., "Road extraction using object oriented classification," *Vis. Tech* **20**(34), 45–50 (2011).
53. M. Belgiu, L. Drăguţ, and J. Strobl, "Quantitative evaluation of variations in rule-based classifications of land cover in urban neighbourhoods using WorldView-2 imagery," *ISPRS J. Photogramm. Remote Sens.* **87**, 205–215 (2014).
54. 2015 Environment for Visualizing Images ENVI, "Environment for visualizing images: version 5.2. Exelis Visual Information Solutions, USA: ITT Industries" (2014).
55. K. J. Danjuma, "Performance evaluation of machine learning algorithms in post-operative life expectancy in the lung cancer," arXiv:1504.04646 (2015).
56. I. RStudio, *"RStudio," Version 1.1.463*, RStudioInc., Boston, Massachusetts (2018).
57. S. Prabhakaran, "Performance analysis and companion functions for binary classification models," InformationValue v1.2.3 (2016).
58. C. Wu, "Normalized spectral mixture analysis for monitoring urban composition using ETM+ imagery," *Remote Sens. Environ.* **93**, 480–492 (2004).
59. G. Xian et al., "Performances of WorldView 3, Sentinel 2, and Landsat 8 data in mapping impervious surface," *Remote Sens. Appl. Soc. Environ.* **15**, 100246 (2019).
60. K. Shahi et al., "A novel spectral index to automatically extract road networks from WorldView-2 satellite imagery," *Egypt. J. Remote Sens. Sp. Sci.* **18**(1), 27–33 (2015).
61. R. Pu and S. Landry, "A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species," *Remote Sens. Environ.* **124**, 516–533 (2012).
62. Y. Zhu et al., "Exploring the potential of World View-2 red-edge band-based vegetation indices for estimation of mangrove leaf area index with machine learning algorithms," *Remote Sens.* **9**(10), 1060 (2017).

Adeniyi Adeyemi received his Bsc (Hon) in environmental management and toxicology from Federal University of Agriculture, Abeokuta, Nigeria and MSc degree in geoinformatics (applied remote sensing and GIS–urban heat island study) from the University of Pretoria, South Africa. He is a PhD candidate at the Department of Environmental science, University of South Africa (UNISA). His research interests include methods for information extraction from remotely sensed imagery and its applications in urban impervious surface studies using multispectral data.

Abel Ramoelo is a remote sensing specialist at the Scientific Services, South African National Parks. He has a PhD in remote sensing from the University of Twente (The Netherlands). His core expertise is on vegetation health/quality and quantity/productivity, biodiversity, and land cover/use mapping using remote sensing systems at various scales. He published over 50 peer-reviewed scientific and conference papers, and is a member of the editorial board for various remote sensing journals.

Moses Cho is the Precision Agriculture research group leader at the Council for Scientific and Industrial Research (CSIR) and a professor with the plant and soil science department,

University of Pretoria. He holds a PhD in remote sensing of vegetation from Wageningen University and the International Institute from Geoinformation Science and Earth Observation, The Netherlands. He has published over 100 journal and conference articles and is an associate editor for the *International Journal of Applied Earth Observation and Geoinformation*.

Cecilia Masemola received her BSc, MSc, and a PhD degrees from the University of South Africa, South Africa. She is a lecturer and geographic information system (GIS) and remote sensing specialist in the Department of Geography at the University of Kwa-Zulu Natal. Her research interests include applying geospatial big data, machine learning, deep learning and cloud computing (e.g., Amazon-Web Services and Google Earth Engine) to study environmental change, agriculture precision and to model land use/land cover dynamics.