Classification for Research Universities in India[1]

Pankaj Jalote
Distinguished Professor (and founding Director)
IIIT-Delhi
New Delhi, 110020
Contact: jalote@iiitd.ac.in, +91 11 2690 7499

Bijendra Nath Jain
Professor, Computer Science, IIT Delhi
(ex-Vice Chancellor, BITS Pilani)

Sudhir Sopory
Professor, Intl. Centre for Genetic Engineering and Biotechnology, New Delhi
(ex-Vice Chancellor, Jawaharlal Nehru University, Delhi)

Abstract. Classification of higher education institutions (HEIs) of a country allows viewing higher education as a differentiated system which respects the diversity of purposes and aspirations of different HEIs. Classification is fundamentally different from ranking, which aims to rank universities in order with higher ranked HEIs being "better" than lower ranked ones. In classification, the universities in a class are grouped by their purpose and mission, and no attempt is made to rank them. Carnegie Classification of universities in the US is the oldest classification system, which groups universities into a few categories like Research Universities, Masters Universities, Baccalaureate Universities, Secondary, etc. This classification has been found extremely useful over decades for various purposes including policy making and planning. This has thus motivated similar exercises in many other countries, particularly for research universities. In this paper, we evolve an approach to classify research universities in India, based on the Carnegie classification approach. We propose a simple basic criterion for identifying research universities, and apply it to the top 100 universities and top 100 engineering institutions in India. Using this criteria 40 universities and 32 engineering institutions were identified as research HEIs. Based on the data on the level of research activity in these HEIs we apply a clustering approach similar to the one Carnegie uses to group research HEIs into two sub-categories, viz. "highest research activity" and "moderate research activity". The clustering approach identified six universities and eight engineering institutions in India to be in the highest research activity category. The level of research activity uses data on the number of full time PhD students, the number of faculty, research grants, and publications.

Keywords. Classification of Higher Education Institutions, Research Universities, Indian Higher Education System.

Introduction

A university has research and higher education as twin focus. However, not all universities emphasise both equally. This gives rise to different types of universities with different overall goals or mission. On the one end, there are teaching focused universities, whose main goal is to provide higher education to students, though they may also engage in research. At the other end of the spectrum are research universities, whose main goal is to create knowledge through research, though they often also pay a lot of attention to their undergraduate and other education programs.

The purpose of classifying universities is to group universities with similar objectives or mission. As Carnegie report states "Classification was designed to support research in higher education by

identifying categories … that would be homogeneous with respect to the functions and characteristics …" (Carnegie 2000). A key goal of classification is to help understand complex systems with a heterogeneous population by grouping entities into sub-groups such that entities in one sub-group share some common features, while differentiating them from entities in other sub-groups (McCormick and Borden 2017). For the higher education system of a country, which is often quite complex, classification helps to capture and describe the diversity in higher education (Carnegie, 2000). It can also help in developing policies for higher education. For example, developing models and policies to support universities based on their mission, or to grant different levels of autonomy to different types of institutions. Classification based on data about universities often helps in formalizing differentiation that may informally exist in their missions, or for reflecting the missions that universities are actually pursuing rather than the claims made (McCormick and Borden 2017). It is also a tool to help consumers make informed choices. For example, a classification for research universities can help a PhD aspirant decide where he/she should seek admission. Some other uses of classification are given in (McCormick and Borden 2017).

Classification is different from university rankings which, by definition, rank order the universities. Most rankings are based on multiple criteria, with different weights assigned to each criterion for obtaining the final score for purpose of ranking. For example, the National Institutional Ranking Framework (NIRF) of India, started by the Government, assigns only 30% weight to research (it assigns 30% to teaching and learning, 20% to graduate outcomes, 10% to outreach, and 10% to perception). Ranking thus reflects a weighted sum of performance in teaching, research, service, perception, etc. This is different from classification, which is to categorize universities based on characteristics they share. The class of research universities will get defined by characteristics relating to primarily to research.

Classification of HEIs is also different from accreditation. In accreditation authorized agencies such as ABET in US and NAAC in India audit processes an HEI uses to manage its operations, based on which an agency will accredit the university at a given level. The process looks at the full range of activities a university is engaged in. For example, the NAAC accreditation framework has multiple assessment criteria, only one of them being research. Others include curricular aspects, teaching-learning pedagogy, infrastructure, governance, student support, etc. Due to its wide scope, NAAC accreditation is also a weighted measure not designed for identifying research institutions. Accreditation is also a voluntary activity in that not all institutions choose to get accredited. Accreditation status of HEIs in India can be found from NAAC website (NAAC Website).

Today, most countries desire, and work towards it, to have some of their best-in-class universities to be ranked amongst the top global universities. Classification can help in identifying universities for such a potential – as pointed out by Altbach "all world class universities are research universities without exception, but all research universities are not world class" (Altbach, 2007). It is clear that if a country wants to have some world class universities, the likely candidates will necessarily have to be those that can be classified as research intensive universities.

Classification is best done at a country level in order to address country-specific missions that HEIs have. As classification is a grouping of HEIs with similar goals/missions with no rank order among them, the criteria for classification should be as simple as possible. The most well-known classification method is the Carnegie Classification, which was started in 1970s for US HEIs. Subsequently, frameworks have been proposed for classification of research universities in China, Korea, EU, Japan, etc. We will discuss these further in the next section. There is currently no classification framework for Indian HEIs.

In India there are about 900 Higher Education Institutions which can grant degrees and more than 40,000 colleges which are teaching institutions affiliated to a University (UGC Website). The NIRF has grouped HEIs into different categories:

- "Universities" (or traditional universities that focus on undergraduate, post-graduate and PhD programs, viz. BA/BSc, MA/MSc, PhD, in various disciplines including Natural Sciences, Humanities and Social Sciences, Management, Law, etc). Some of them may also offer programs in Engineering disciplines.
- "Engineering" institutions (or degree granting HEIs that have a strong focus on Engineering, but often also have sciences, management and a few other disciplines.).
- Specialized degree granting HEIs that focus mostly on one discipline such as Management, Medicine, Pharmacy, Architecture, or Law.
- "Colleges", that focus mostly on undergraduate education and do not have degree granting powers. A college delivers programs designed by an affiliating university, which also undertakes assessment and grant of degrees. College do not have PhD programs.

For the classification work reported here, we have considered the two largest categories of HEIs, viz. "Universities" and "Engineering" only, and have not considered "Specialized HEIs" nor "Colleges" (Colleges are not considered since research is not in their scope). These two categories, viz. Universities and Engineering institutions, cover almost all the well-known HEIs in India – and are sufficiently broad in scope to allow one to define what constitutes a research university. Also, HEIs that specialize in a single discipline, for instance, Law, Management, Medicine, etc., will require specialized criteria for research as they are often more practice oriented. It may also be added that only about 10% of the students are enrolled in the specialized programs like Medicine, Law, Management, etc. (UGC website). More information about the ranking approach can be found in (NIRF 2015, Varghese 2018).

These two types of HEIs – Universities and Engineering Institutions –not only have the largest number of HEIs, they are also the two main categories from governance perspective in India – Universities generally have a Vice Chancellor as the Chief Executive while Engineering Institutions have a Director as the Chief Executive – the role and power of the two are somewhat different. The academic programs also are often different– Universities generally focus on offering 3-year Bachelor programs in Natural Sciences, Social Sciences, Humanities, etc, while Engineering HEIs predominantly offer 4-year BTech or BE degrees. From an Indian perspective, these two are the main categories, and are often considered quite distinct, with different regulating bodies for them: UGC (Universities Grants Commission) for universities and AICTE (All India Council for Technical Education) for engineering institutions.

The NIRF site provides data for the 100 top HEIs in each of these two categories (for its 2018 exercise). As the number of HEIs that can be considered as research universities is likely to be relatively small, we believe that considering the 100 top HEIs in each category is sufficient for this classification. (We observed that there are six universities that are not in the list of top universities but are included in the top "Overall", which uses a different criteria. We have included these six also in Universities for our analysis).

For classifying research HEIs in the two categories, we follow the two-step approach that Carnegie follows. We use simple basic criteria to separate out Research HEIs from the rest. Then we use research activity measures and apply a clustering technique to sub-classify research HEIs in two groups – ones with highest research activity, and those with modest research activity.

The rest of the paper is organized as follows. In the next section we briefly discuss different approaches to classification used in US and elsewhere. We then describe the approach we use for India, and also highlight some specific circumstances of Indian HEI scenario. We then present the results of applying the framework to the top HEIs in the two categories of HEIs, followed by our conclusions.

Research University Classification Frameworks

Carnegie Classification is the oldest and most influential classification framework. Started in 1970, it classifies HEIs into a few broad categories: Doctoral/Research Universities, Masters Colleges and Universities, Baccalaureate Colleges, Associate Colleges, Specialized Institutions, and Tribal Colleges and Universities. Of a total of over 4500 HEIs considered in the2015 classification, the number of Research Universities is about 7% of the total.

For classifying research universities, a two-stage process is used. A simple basic criteria for a Research University is used to separate research universities from the rest – a university is defined as a Research University (RU) if it has graduated more than 20 PhDs per year in the recent past (in an earlier classification, this number was 50 PhDs per year). Based on this basic criterion, 335 universities are classified as RUs in the 2015 edition.

The basic classification separates research universities from the rest. However, this class itself contains a range of universities. For example, this set of research universities includes universities such as MIT, Caltech, UC Berkeley, UIUC, GaTech, CMU, etc. where the number of PhDs graduated *per faculty* per year is 0.5 or higher, and where sponsored research is in 100s of million dollars, as well as many universities where the number of PhDs graduated per faculty per year is less than one tenth of this. Hence, these are further sub-classified.

In the second stage of classification, the RUs are grouped into three sub-categories: R1 (highest research activity), R2 (higher), and R3 (moderate). The following features related to their research activity are considered while grouping the RUs into the three sub-categories, viz. R1, R2 and R3:

- Number of faculty members,
- Research manpower,
- Number of PhDs granted, and
- Research funding.

These features are considered to be the most defining features of a research university and, therefore, used for the purpose of classification. In addition to research faculty, an RU also requires research manpower. Hence, this factor is included. Globally, the main research manpower (besides faculty) is the PhD students. In advanced countries such as the US, however, RUs also employ a considerable number of post-doctoral staff for research. In Carnegie Classification post-doctoral fellows are counted as research manpower.

A fundamental difference between an RU and a teaching-focused institution is the size and importance of the PhD program in the RU. In fact, Carnegie Classification considers this feature *only* for basic classification of a university as an RU. For sub-classification it considers number of PhDs granted in STEM and HSS fields.

Clearly, funding is needed to conduct research, including funds to support PhD students or employ research staff as also to develop and maintain lab equipment. Globally, while universities do provide limited support for research, much of financial support for research comes in the form of externally-sponsored research grants. Hence, an RU seeks funding for research projects to partly pay for its research manpower and research equipment and facilities. Thus, the amount of research funding is a strong indicator of research activity. In Carnegie, this is called R&D expenditure in STEM and non-STEM areas.

It should be pointed out that for purpose of classification, the focus is on a few key parameters that capture the level of research activity. Qualitative assessment (e.g. the quality or impact of research), which may be important for ranking, is generally not considered in classification.

For grouping into the sub-categories, Carnegie does a clustering analysis using these features to group

them into three sub-categories. The clustering approach first groups the features into two sub groups – aggregate (i.e. the total value), and per-capita (i.e. features normalized by faculty strength). Then, a Principal Component Analysis (PCA) is performed for each of the two feature groups to identify the principal component giving (as normalized value) the aggregate-research-index and per-capita-research-index. (PCA analysis is a technique used to reduce dimensionality, because of which it loses some information and cannot, therefore, account for *all* the variability in the data. In the Carnegie analysis, the first principal component accounted for about 70% of the variability in the data.) The values of these two indices for each university are used to have a scatter plot of the 335 RUs. These two values for RUs are also used for clustering the RUs in the three sub-categories (the algorithm used for clustering is not specified). Based on the clustering, they have identified three sub-categories, termed R1, R2, and R3, each with approximately one-third of the 335 RUs earlier identified. More discussion about the methodology can be found in (Kosar and Scott 2018); some ideas behind the Carnegie Classification framework and challenges it faces are discussed in (McCormick and Zhao 2005).

While Carnegie classification is the oldest and the most influential, there have been attempts in other countries such as China, Japan, Korea, and Australia for classifying universities as research universities. Most of these efforts have been influenced by the Carnegie classification. Some of these are briefly discussed here (we could not find an English language reference for Japanese classification work). EU has evolved a somewhat different framework for classification of RUs.

A two-step process for separating Research Universities was undertaken to classify Korean universities (Shin 2009). For basic classification, the criteria used was (a) the "number of PhDs produced is more than 20 per year", and (b) the "number of papers published each year in indexed journals is more than 100". Using these basic criteria, 47 universities were identified. These were then grouped into different categories using a hierarchical clustering approach using key parameters such as faculty size, publications, research funding, and PhD students graduated – the last three performance parameters being normalized with respect to faculty size. As a result, the universities were grouped into five clusters based on their research performance.

In the Chinese classification framework, four features were used (Liu 2006, Liu 2007). These are: (a) total number of degrees awarded at different levels, (b) ratio between doctoral and baccalaureate students, (c) annual research income, and (d) per capita research articles in indexed journals. The universities are classified into a few different categories, with Research Universities being grouped into two sub-categories: Research Universities I (7 universities), Research Universities II (48 universities).

Corresponding work on Australian universities is more about identifying quantitative performance indicators that can predict the university type, where the types are pre-defined based on the evolution of Australian universities – Sandstone Universities, Universities of Technology, Wannabee sandstones, New Universities (Ramsden 1999). Initially, nine parameters are considered as performance indicators. Later, this was simplified by forming two constructs based on fewer variables, one of which constructs considers percentage of staff with PhD, student-staff ratio, and students going for further study.

The EU classification framework also aims to map the characteristics of universities to capture their diversity (Vught 2010). The final outcome, however, is different from the approaches used by Carnegie or from the work done in other countries as discussed above. It does not group universities into a set of labelled categories. It instead categorises them for a range of different characteristics. For mapping different characteristics, they have identified six dimensions: (a) teaching and learning profile, (b) student profile, (c) research involvement, (d) involvement in knowledge exchange, (e) international orientation, and (f) regional engagement. For each of these dimensions a few indicators are identified, with a total of 23 indicators. Based on the data for universities, they are grouped for each indicator into categories such as: major, substantial, some, none; small, medium, large, very

large; etc. This type of classification across multiple dimensions allows universities to determine similarities and dissimilarities among each other along these dimensions.

A different approach for classification has been proposed for humanities and social sciences departments, in which departments are sought to be classified in three worlds – the top tier (elite), the middle tier (pluralist), and low tier (communitarian) (Hermanowicz 2005). This classification focuses on departments rather than institutions and is based on the culture of the faculty in terms of how they view their role and their professional careers progression, which is believed to be defined largely by the organizational framework.

Methodology for Classifying Research Universities in India

For classifying research HEIs in India, as in the Carnegie framework, we also propose a two-step approach: (a) a simple basic criteria to separate research HEIs from the rest, and then (b) a more involved sub-classification by clustering research universities identified in the first step using data on their levels of research activity. In this section we describe both of these. Before doing so, we discuss some aspects of the Indian higher education system which are considered important in the country, and which need to be considered while defining the classification criteria.

Higher Education System in India

In India, the higher education system has grown very differently from the way it has in the US (or in other countries such as Australia, UK). Instead of broad-based universities with multiple schools and departments, it has grown by having HEIs that are focused on a few disciplines. Consequently, most HEIs tend to be smaller as compared to their global counterparts. For example, more than half of the HEIs have student strength of less than 5000, while a vast majority of the top research universities in the world have a student population of more than 10,000. Hence, any framework for research HEIs in India should account for the fact that most HEIs will be modest in size.

Carnegie, and some other classification approaches, assume implicitly that all or most faculty in universities hold doctorates. In India, that is not the case - there are a large number of HEIs that have many faculty members who do not have doctorates. For this reason NIRF collects data separately on the total number of faculty members that have a PhD, and those that do not. Consequently and necessarily, in order to identify research HEIs, we consider the total faculty strength, and the ratio of faculty members who have a PhD.

A fundamental difference between a research HEI and teaching-focused institution is the size and importance of its PhD program. In fact, Carnegie considers this feature alone for classifying a HEI as a research HEI, or otherwise. In India, since focus on research in many universities is a recent phenomenon, and many of the HEIs that are focused on research have been created only in this century, we feel that for such a growing system it is better to capture the strength of the PhD program in terms of the total full-time PhD student population (rather that number of PhDs graduated in last one year). Further, since almost all full-time PhD students in India receive some form of scholarship, the number of full-time PhD students enrolled is a strong indicator of research activity as well as research investment. (Note also that in the steady state this criterion can be easily converted to number of PhDs graduated.)

For research manpower, the Carnegie approach considers post-doctoral fellows as the research manpower (besides faculty). In India, there is literally no tradition of employing post-doctoral fellows – even the better known institutions for research, viz. IITs, hardly have any. Hence, for research manpower, we focus on PhD students (besides faculty). Other classification frameworks have also considered PhD students as the primary research manpower.

Criteria for Basic Classification

Clearly an HEI that is focused on research must have research faculty. The world over, research faculty predominantly hold doctorates. In fact, a hallmark of research universities is that they mostly

employ as full time faculty those that hold PhDs (Altbach 2007). Given that a large fraction of faculty in many HEIs in India do not possess a PhD, we require that at least 75% of the faculty have doctorates before an HEI qualifies to be considered as a research HEI.

A reasonable expectation for a research HEI is that each faculty member has on an average one full time PhD student working with him/her. This should be the case for a research HEI regardless of whether it has a focus on social sciences, physical sciences, engineering or any other discipline and hence is quite general and can be applied to both the categories of HEIs we are considering. We use this as the basic criterion for defining a research HEI in India. For the purpose of this study, we assume that all full-time PhD students are paid stipend or fellowship at a level approved by the regulator (i.e. UGC, AICTE) or the Government.

With this, the basic criteria for an HEI to qualify as a Research HEI in India is:
> RU-C1: % of faculty with PhD > 75% of total faculty, and
> RU-C2: Ratio of number of full time PhD students to number of faculty is > 1.

This basic criterion can be applied to different types of HEIs, and is similar in spirit to the basic criteria used by Carnegie in that it focuses on PhD students – except that we have added an additional test on percentage of faculty with PhD – a test necessary for HEIs in India. (Also, while not explicitly stated, we assume that a research HEI has more than 50 faculty members – this holds true for all the HEIs we have considered.) Such a criteria can also be easily extended later to define other categories of HEIs also – e.g. for Masters HEIs, as in Carnegie. We may, for instance, later suggest that an HEI is categorized as a Masters HEI if the ratio of PhD students and faculty is < 1.0, but the ratio of Masters students and the faculty size is greater than some threshold.

Approach for Sub-Classification of Research Universities
For *sub-classification* of research HEIs using clustering, the main features we consider are:
1. amount of sponsored research grants (similar to Carnegie's research expenditure which they divide in two categories – STEM and non-STEM spending),
2. the total number of full time PhD students (Carnegie considers total no of PhDs granted, which they split into four categories),
3. the total number of faculty, and
4. the total number of publications in indexed journals.

It may be noted that Carnegie classification does not include publications in its methodology, but it is an important parameter that distinguishes more active research universities from the less active ones. The Chinese and Korean classification approaches also consider publications in indexed journals.

In Carnegie, for clustering research universities into different sub-groups, as mentioned above, they define Per Capita Research Activity Index and Aggregate Research Activity Index based on the value of the key research features of the university. With these two indices, the universities are plotted on a 2-dimensional plot, and a clustering approach is used to cluster them into three clusters – R1, R2, and R3, representing (i) highest research activity, (ii) high research activity, and (iii) moderate research activity sub-groups. Carnegie does not specify the clustering algorithm they have used.

We also consider two feature-sets – one is aggregate, and the other is normalized by the number of faculty. We also do a PCA to identify the main principal component for both the aggregate and normalized feature-sets, and then use the extracted aggregate research activity index and normalized research activity index to plot them and cluster them.

For clustering, we use the standard k-means algorithm (Duda 2000). Given that our data set is rather small (we have less than 50 for each of the two types of HEIs, as compared to 300+ which Carnegie had), we decided that separating them in two clusters is more meaningful – R1 and R2. With two clusters, R1 will represent the HEIs with highest research activity, and R2 will represent those with modest research activity.

It should be pointed out that in the K-means approach, the clustering is done completely algorithmically, and the analyst provides no input parameters other than the number of desired clusters. This helps make this approach also neutral and minimizes bias or subjectivity on the part of the analyst, if any.

Classification Analysis Results

Getting accurate data is, of course, critical for any classification (or even ranking) exercise. Many countries have some government agency collecting data for policy making purposes. In India the National Institutional Ranking Framework (NIRF 2015) - launched three years ago - has been widely accepted in academic circles. As NIRF is sponsored by the Government agency, it is believed that the data provided by the HEIs to NIRF is more likely to be complete, and the checks done by NIRF more rigorous. We feel that this may be the most accurate and reliable data available in India.

NIRF has published data for top 100 HEIs in most categories on its website. Of course, as NIRF is a ranking agency, it compiles a lot of data, including data on placements, learning outcomes, etc. For this work, we prepared a database of the required data about the institutions from the NIRF site for the year 2018 (which has data for 2017). In this paper, the data we have used is exclusively from the information reported on NIRF website about the institutions.

To obtain the data of the top HEIs in each of the types of HEIs, we downloaded the public data tables for each HEI published by NIRF on its website, given in *pdf* format. Of course this data is much more than what is needed for classification. We extracted the data we need from this *pdf* document through a script (a separate one for each NIRF institution type). Specifically, we extract data on number of faculty, number of faculty with PhD, number of full time PhD students, research grants, and publications (Scopus indexed) – the attributes we need for classification. (To verify we manually checked the data extracted for about quarter of the HEIs.)

It is to be noted that 24 institutions are listed in the top 100 institutions in both the types of HEIs that we are considering – Universities and Engineering. That is, they are listed as a University as well as an Engineering institution – these 24 institutions are mostly broad-based universities which have Engineering Colleges, and hence are included in both. Some Engineering Institutions may be included by NIRF in the University category only because they were created by a University Act of an individual state, and thus designated as a State University. For these 24 institutions, NIRF has collected separate data for them as a University as well as an Engineering institution, with the data on it as an Engineering institution pertaining only to the engineering programs. For our analysis we consider the two groups separately, and these common institutions, are considered in each group.

Basic Classification

We applied this criterion to the top HEIs in the two categories of NIRF, as discussed above. As a result, the number of HEIs from the two groups that can be classified as RUs is given below:

Category of HEI (as per NIRF)	Total no of HEIs considered	No of Research HEIs
University	106	40
Engineering	100	32

The total number of HEIs that satisfy the basic criteria is 68 – with 4 of these listed in both categories. (It should be pointed out that of the HEIs listed in both, there is none which satisfied the RU criteria in one but not the other – all of them either satisfied the criteria in both or did not do so in either category.)

This number of RUs (68, that is) also seems reasonable – most academics in India will agree that the total number of HEIs that can be considered as research HEIs is definitely not very large. It is also not inconsistent with the general pattern that the number of research universities is likely to be less than 10% of the total number of HEIs. The number is also comparable to the number of research universities in China and Korea (as per their classification).

The list of HEIs in the two types of institutions that satisfy the criteria, along with relevant data on total number of faculty, number of faculty with PhD, and the number of full-time PhD students, are given in Table 1 and Table 2 in Appendix 1.

Of the HEIs in each of the two categories that did not satisfy the criteria to be classified as a research HEI, vast majority did not satisfy both the components of the criteria (percent of faculty with PhD > 75%, and number of FT PhD students > number of faculty), though there were some which did not satisfy one or the other basic criteria. Of the 66 Universities that were not classified as research HEIs, 42 did not satisfy both criteria. Of the 68 Engineering HEIs, 56 did not satisfy both conditions. Only a few HEIs satisfy one criteria and not the other.

It is interesting to note that, in the top 25 engineering institutions in NIRF ranking, there are six that do not qualify as research universities. Some private institutions (e.g. BITS Pilani, Thapar, VIT), which are known for their good quality of education and are reputed, do not satisfy the criteria for research universities - in fact both the conditions for research university are not met. Similarly, in the top 25 universities in NIRF ranking, there are fourteen, many of them private, that do not satisfy the criteria for a research university. This clearly shows that in a ranking framework like NIRF which places strong emphasis on UG education, placement of its graduates, etc., some institutions that are considered good in education and have a long reputation may be ranked high, but which may not satisfy the criteria for being classified as a research universities since they are not focused on research.

It is also worth noting that all the HEIs that satisfy the criteria for a research university are public institutions – 23 Universities and 28 Engineering Institutes are centrally funded, rest are funded by state government (or a combination of state and centre). Only one institution is classified as private, but it was created by a state government, which also funded its initial infrastructure and development). This is mostly due to the fact that private institutions are self-supporting and depend solely on revenue from tuition and other student fees. Consequently, they are not able to support research at any reasonable level, nor provide for at least one full-time PhD student per faculty. It is worth pointing out that private institutions are sometimes not eligible for research grants from some research funding agencies, making it harder for such institutions to support research.

Sub Classification of Research HEIs
As mentioned above, we also use the PCA analysis on the features to define the aggregate research activity index, and the normalized (by the number of faculty) research activity index. (In our analysis, as in Carnegie, the main Principal components account for about 70% of the variance.) We used these two indices for clustering, using the standard k-means algorithm (Duda 2000). We clustered them into two clusters – R1 and R2. One may conclude R1 represents the HEIs with highest research activity, and R2 represents those with modest research activity.

For the 40 Universities that satisfy the research criteria, the clustering approach identified six universities with the highest research activity. The scatter plot for the 40 Universities is given in Figure 1. (Had we considered grouping these into 3 clusters, Indian Institute of Science would have showed up in a cluster of its own, and the others from R1 showed up in the second cluster.)
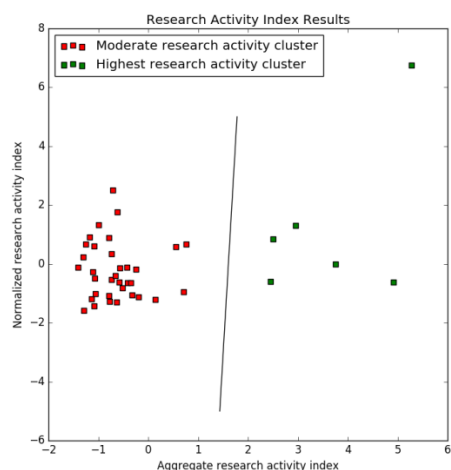
Figure 1: Plot and clustering for research universities in India

The list of universities that fall in R1 (highest research activity index) along with the value of their normalized features, i.e. number of full time PhD students per faculty, number of *Scopus* indexed publications per faculty, and research funding (in INR 100,000) per faculty, are given in Table 3 (in alphabetical order). (Their values for total number of faculty, number of faculty with PhD, number of FT PhD students are given earlier in Table 1).

Table 3. List of highest research activity universities (in alphabetical order).

University	Total Faculty	PhD students /Faculty	Research Funding /Faculty	Publications / Faculty
Banaras Hindu University	1619	2.2	9.1	2.7
Homi Bhabha National Institute	1014	1.7	24.6	0.5
Indian Institute of Science	430	6.2	91	17.4
Jadavpur University	643	4.1	8.6	7.3
Jawaharlal Nehru University	652	8.3	7	3.6
University of Delhi	1055	3.1	5.7	5.1

For the 32 Engineering Institutions that satisfy the research HEI criteria, on applying this approach, a total of eight HEIs were included in R1. The scatter plot for these is shown in Figure 2.
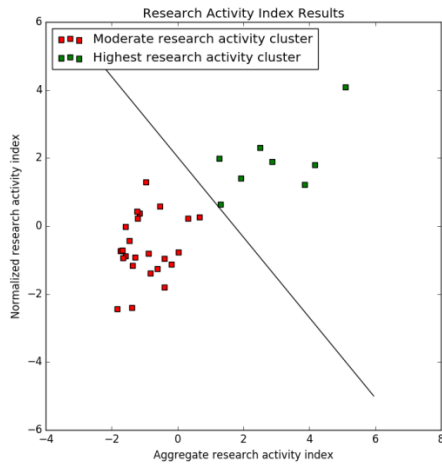
Figure 2: Plot and clustering of Research HEIs (Engineering) in India

The list of Research HEIs (Engineering) that are in R1 (highest research activity), along with the values of the normalized features, i.e. number of full time PhD students per faculty, number of Scopus indexed publications per faculty, and research funding (in INR 100,000) per faculty, are given in Table 4 (in alphabetical order). (Their values for total number of faculty, total number of faculty with PhD, number of FT PhD students are given earlier in Table 2).

Table 4. List of highest research activity Engineering HEIs (in alphabetical order).

Institution	Total Faculty	PhD students /Faculty	Research Funding /Faculty	Publications / Faculty
Indian Institute of Technology Bombay	528	4.1	65	7.8
Indian Institute of Technology Delhi	481	3.5	19.5	8.2
Indian Institute of Technology Guwahati	401	3.4	8.5	5.6
Indian Institute of Technology Kanpur	418	3.6	38.5	6.6
Indian Institute of Technology Kharagpur	644	3.6	11.3	6.8
Indian Institute of Technology Madras	607	3.3	32.1	6.6
Indian Institute of Technology Roorkee	423	3.7	7.2	7.7
Jadavpur University	323	3.7	9.9	9.3

(Jadavpur University is included both in Universities and Engineering HEIs. It is classified as a research HEI under both categories, and as it turns out, it is included in R1 under both these categories. The value of key features in the two is different – in Table 4, NIRF data for Jadavpur University is limited to its Engineering College and related departments only.)

The list of Universities and Engineering Institutions that fall in the R1 category contains HEIs that are widely respected and recognized for their faculty quality and academics. And most academicians will agree that these are indeed the best universities/engineering institutions in terms of research in the country. Whether some other HEIs should also be considered part of the R1 sub-group is a matter of opinion and arguments can be made in favour of some. However, the clustering done above is done algorithmically with no guidance to the algorithm. Also, we can see that the grouping is visually quite evident – there is a clear separation between the two groups.

Conclusion

In contrast to ranking, classification of universities groups universities into a few categories, depending on their mission and goals. Carnegie classification of universities in USA is the oldest and most influential classification scheme. It classifies universities into seven categories, one of them being research universities. They use simple basic criterion for classifying a university as a research university, viz. the number of PhD students graduated. It further sub-classifies research universities into three sub-categories: R1 (highest research), R2 (high research), and R3 (moderate research) by clustering them based on the aggregate level of research and per per-capita level of research.

In this article, we have evolved a classification framework for research HEIs for India, based on the Carnegie framework. For separating the research HEIs from the rest, we have used the criteria that 75% of the faculty has a PhD, and the ratio of full-time PhD students to faculty is more than one. To further sub-classify the research universities, we determined the aggregate research activity index and normalized research activity index, and then used the K-means clustering approach to identify the "highest research activity" HEIs and the "moderate research activity" HEIs.

By applying the basic criteria, we found that 40 universities and 32 Engineering institutions meet the criteria to be grouped as research HEI. This constitutes about 7% of the degree granting HEIs in India, which is somewhat similar to what Carnegie's classification has reported for the US. And the total number of research HEIs is similar to those in China and Korea, as per their classification scheme.

Of the research HEIs, we found that 6 universities and 8 Engineering Institutions come under the category of "highest research activity", and the clustering chart shows clear separation of this group of HEIs with the rest of the research universities/engineering institutions. For universities this is about 20% of the research universities, and for Engineering HEIs about 25% of the research institutions. This is somewhat lower than about one-third which Carnegie classifies as "highest research activity" universities within the research universities. While HEIs that are included in R1 are widely recognized for their research, whether or not some other HEIs should also be included in R1 sub-group is a matter that may be argued in favour or against inclusion. HEIs at the boundary of the two clusters may be considered for inclusion in sub-group R1 by providing guidance to the clustering algorithm regarding the size of the clusters or other constraints. We have not done any of this presently, and have relied entirely on the standard k-means algorithm for clustering.

We feel that the universities in R1 category have a high potential to make it to world rankings, particularly if their size and scope, as well as funding levels, are expanded to global levels. In fact, in some world rankings in certain years (e.g. QS, 2018) institutions such as IISc, IIT Bombay, IIT Delhi are already in the top 200.

To strengthen research in universities so that some of them reach global rankings, India will need to identify and support a reasonable number of research universities – it is unrealistic to have all universities focus on research in a large system like that of India, where resources are also very limited. While the top few universities are easy to identify in India, and this classification has also identified them, if more universities are to be supported to strengthen research in the country, a better understanding of research universities will be needed. Classification approach like the one presented here can help in this. For example, from the R2 group, universities can be critically examined to identify their weaknesses and potentials, and be supported so they can move to R1 over time. It can also help universities in understanding their current research level, and develop plans for moving from R2 to R1 - as has been the case in the US. The basic classification can help those universities that aspire to be research focused identify necessary steps for the same. Clustering can also help in formulating criteria on the input parameters for sub-grouping of RUs.

In the work reported here, we have used the data collected by NIRF and reported on its website, and have considered data for one year, which was appropriate as NIRF itself is very young. However, after a few years, average of previous two or three years data can be used for this classification.

We feel that this is an initial exercise to define criteria and methodology for identifying research HEIs in India. With further discussion and research, the approach can be refined further, and with time, as the higher education system in India matures and reaches some level of stability, the criteria can be suitably enhanced. We also feel that such a classification should be done every few years to understand the evolution of research universities in India. This has been done by Carnegie also, and will be particularly useful for India as the higher education system is evolving and expanding rapidly. The approach presented here can also be expanded in future to cover the specialized HEIs also.

Further work is also needed to expand this approach for identifying other types of HEIs and evolve a comprehensive framework, like the Carnegie, for classifying HEIs, in multiple categories, including research HEIs.

References

1. Altbach, Philip G (2007). Empires of Knowledge and Development, in *World Class Worldwide*, eds: Philip G Altbach and Jorge Balan, Johns Hopkins Press, 2007.
2. Carnegie (2000), The Carnegie Classification of Institutions of Higher Education, http://carnegieclassifications.iu.edu/downloads/2000_edition_data_printable.pdf
3. Carnegie (2016). 2015 Update – Facts and Figures, http://carnegieclassifications.iu.edu/downloads/CCIHE2015-FactsFigures-01Feb16.pdf
4. Duda, Richard O., Hart, Peter E., and Stork David G. (2000).*Pattern Classification*, 2nd Ed., Wiley, 2000.
5. Hermanowicz, Joseph C (2005). Classifying Universities and Their Departments: A Social World Perspective, *The Journal of Higher Education*, Jan/Feb 2005: 26-55.
6. Liu, Nian C. (2007). Research Universities in China, in *World Class Worldwide*, eds: Philip G Altbach and Jorge Balan, Johns Hopkins Press.
7. Liu, Nian C. (2006). Classification of Chinese Higher Education Institutions, online on oecd.org
8. NAAC website. https://assessmentonline.naac.gov.in/public/index.php/hei_dashboard
9. NIRF (2015). A Methodology for Ranking of Universities and Colleges in India, 2015, https://www.nirfindia.org/Docs/Ranking%20Framework%20for%20Universities%20and%20Colleges.pdf
10. Kosar, Robert & Scott, David W. (2018). Examining the Carnegie Classification Methodology for Research Universities, *Statistics and Public Policy*, 5:1, 1-12.
11. McCormick, A. C., and Zhao, Chun-Mei (2005). Rethinking and reframing the Carnegie Classification, Change, Sept/Oct 2005.
12. McCormick, A. C., and Borden, V. M. H. (2017). Higher education institutions, types and classifications of. In J.C. Shin, P. Teixeira (eds.), Encyclopaedia of International Higher Education Systems and Institutions, https://doi.org/10.1007/978-94-017-9553-1_22-1
13. Ramsden, Paul (1999). Predicting institutional research performance from published indicators: A test of a classification of Australian university types, *Higher Education*, 37: 341-358.
14. Shin, Jung C. (2009). Classifying higher education institutions in Korea: a performance-based approach, *Higher Education*, 57: 247-266.
15. UGC Website. https://www.ugc.ac.in/stats.aspx
16. Varghese, N. V. (2018). The New National Rankings in India, *International Higher Education*, Number 93: spring 2018
17. Vught, van, Kaiser,F.A., File, F., Gaethgens, J. M., Peter, C.,Westerheijden, R.(2010).U-Map: The European Classification of Higher Education Institutions, http://www.u-map.eu/U-MAP_report.pdf

Appendix. HEIs identified as Research HEIs from the basic criteria

Table 1: List of Research Universities (in NIRF order)

Institute Name	City	#Faculty with Ph.D	#Total faculty	#FT PhD students
Indian Institute of Science	Bengaluru	430	430	2681
Jawaharlal Nehru University	New Delhi	593	652	5432
Banaras Hindu University	Varanasi	1228	1619	3553
University of Hyderabad	Hyderabad	377	402	1584
Jadavpur University	Kolkata	573	643	2613
University of Delhi	Delhi	827	1055	3293
Jamia Millia Islamia	New Delhi	541	689	1350
Bharathiar University	Coimbatore	273	294	346
University of Madras	Chennai	253	263	772
Institute of Chemical Technology	Mumbai	109	116	543
Andhra University	Visakhapatnam	493	580	838
Homi Bhabha National Institute	Mumbai	888	1014	1738
Alagappa University	Karaikudi	251	290	474
Tezpur University	Tezpur	223	290	398
Kerala University	Thiruvananthapuram	187	242	1144
Tata Institute of Social Sciences	Mumbai	266	333	1055
Mahatma Gandhi University	Kottayam	93	115	464
Guwahati University	Guwahati	285	349	1092
University of Kashmir	Srinagar	390	467	898
University of Jammu	Jammu Tawi	284	376	623
Madurai Kamaraj University	Madurai	222	234	510
Pondicherry University	Puducherry	358	380	2428
North Eastern Hill University	Shillong	282	331	931
Bharathidasan University	Tiruchirappalli	190	242	442
Cochin University of Science and Technology	Cochin	133	151	545
Calicut University	Malappuram	135	145	608
Bidhan Chandra Krishi Vishwavidyalaya	Nadia	220	240	274
Maharshi Dayanand University	Rohtak	337	396	783
The Gandhigram Rural Institute	Gandhigram	155	203	290
Mizoram University	Aizwal	160	208	551
Kalyani University	Kalyani	158	179	705

Assam University	Silchar	302	328	491
Periyar University	Salem	147	161	418
Nagaland University	Zunheboto	160	194	248
International Institute of Information Technology	Hyderabad	77	85	118
Indian Institute of Science Education and Research Pune	Pune	147	147	453
Indian Institute of Science Education & Research Mohali	Mohali	94	94	454
Indian Institute of Science Education & Research Bhopal	Bhopal	89	89	278
Indian Institute of Science Education & Research Thiruvananthapuram	Thiruvananthpuram	80	81	184
Indian Institute of Science Education and Research Kolkatta	Mohanpur	105	105	382

Table 2: List of Research Engineering HEIs (in NIRF order)

Institute Name	City	#Faculty with Ph.D	#Total faculty	#FT PhD students
Indian Institute of Technology Madras	Chennai	606	607	1975
Indian Institute of Technology Bombay	Mumbai	527	528	2190
Indian Institute of Technology Delhi	New Delhi	476	481	1690
Indian Institute of Technology Kharagpur	Kharagpur	641	644	2321
Indian Institute of Technology Kanpur	Kanpur	418	418	1525
Indian Institute of Technology Roorkee	Roorkee	419	423	1564
Indian Institute of Technology Guwahati	Guwahati	388	401	1368
Indian Institute of Technology Hyderabad	Hyderabad	181	183	525
Institute of Chemical Technology	Mumbai	108	116	332
Jadavpur University	Kolkata	287	323	1186
Indian Institute of Technology (Indian School of Mines) Dhanbad	Dhanbad	311	340	1102
Indian Institute of Technology Indore	Indore	116	116	370
National Institute of Technology Rourkela	Rourkela	278	299	806
Indian Institute of Technology Bhubaneswar	Bhubaneswar	129	129	211
Indian Institute of Technology (Banaras Hindu University) Varanasi	Varanasi	298	316	801
National Institute of Technology Surathkal	Surathkal	246	302	574
Indian Institute of Technology Ropar	Rupnagar	115	115	224
Indian Institute of Technology Patna	Patna	110	110	317
National Institute of Technology Warangal	Warangal	238	309	451

Indian Institute of Technology Gandhinagar	Gandhinagar	64	65	145
Indian Institute of Engineering Science and Technology Shibpur	Howrah	227	254	413
Visvesvaraya National Institute of Technology	Nagpur	196	234	328
Jamia Millia Islamia	New Delhi	101	102	251
International Institute of Information Technology	Hyderabad	77	85	118
National Institute of Industrial Engineering	Mumbai	59	59	98
National Institute of Technology Durgapur	Durgapur	165	179	293
Motilal Nehru National Institute of Technology	Allahabad	191	215	382
Indian Institute of Technology Jodhpur	Jodhpur	54	54	136
AU College of Engineering	Visakhapatnam	146	153	208
Indraprastha Institute of Information Technology Delhi	New Delhi	64	71	119
Indian Institute of Information Technology Allahabad	Allahabad	63	63	127
Pandit Dwarka Prasad Mishra Indian Institute of Information Technology Design and Manufacturing (IIITDM) Jabalpur	Jabalpur	68	68	91