ISSN : 2210-142X

International Journal of Computing and Digital Systems (IJCDS)

Volume No. 13
Issue No. 2
May - August 2024

ENRICHED PUBLICATIONS PVT. LTD

JE-18, Gupta Colony, Khirki, Extn, Malviya Nagar, New Delhi-110017
PHONE : - + 91-8877340707
E-Mail : info@enrichedpublications.com

ISSN : 2210-142X

International Journal of Computing and Digital Systems (IJCDS)

Aims and Scope

International Journal of Computing and Digital Systems (IJCDS) is a peer-reviewed International Journal that currently publishes 6 issues annually. IJCDS journal publishes technical papers, as well as review articles and surveys, describing recent research and development work that covers all areas of computer science, information systems, and computer / electrical engineering.

The topics covered by IJCDS are including and not limited to the following research areas:

- Reconfigurable Computing & Embedded systems
- Computer Communications and Networking
- Internet of Things & Real Time Systems
- Cyber Security
- Cloud Computing
- Smart Systems
- Information Systems and Communication Service
- Innovation/Technology Management
- Business Information Systems
- Software Engineering
- Mobile & Web Applications
- Theory of Computation
- Data Structures, Cryptology and Information Theory
- Artificial Intelligence & Robotics
- Image Processing, Computer Vision, Pattern Recognition & Graphics
- Data Mining & Big Data
- Smart Grids & Renewable Energy
- Human Computer Interaction

Responsiveness Speed

Time from submission to first decision after peer review: 4-8 Weeks
Time to immediate reject: 2-4 Weeks

Publication Speed

Time from final acceptance to print: 3-6 Months
Total time to publication: 4-8 Months
Time from submission to acceptance: 4-12 Weeks
Time from acceptance to online Publishing: 4-8 Weeks

Publishing Ethics

Plagiarism in all its forms constitutes unethical publishing behavior and is unacceptable. The journal has no tolerance on plagiarism. All submitted manuscripts must go through cross checking using turnitin as an online plagiarism checker.

ISSN : 2210-142X

International Journal of Computing and Digital Systems (IJCDS)

(Volume No. 13, Issue No. 2, May - August 2024)

Contents

ISSN 2210-142X

Performance Evaluation of Incremental Conductance and Adaptive HCS MPPT Algorithms for WECS

Ahmed Badawi1, Hassan Ali1, I. M. Elzein1 and Alhareth M. Zyoud2
1College of Engineering and Technology, University of Doha for Science and Technology, Doha, Qatar
2Department of Electrical and Computer Engineering, Faculty of Engineering, Birzeit University, Birzeit, Ramallah, Palestine

ABSTRACT

This paper presents a novel Maximum Power Point Tracking (MPPT) algorithm designed for Wind Energy Conversion Systems (WECS) to achieve optimal power extraction (Pmax). The controllers employed in this study utilize a Direct Power Control (DPC) framework to assess efficiency and performance, particularly under uncertain and rapid variations in wind speed profiles. The research aims to evaluate the effectiveness of the Incremental Conductance (INC) and adaptive Hill-Climbing Search (HCS) algorithms for MPPT in WECS under such conditions. The modeling of the WECS system utilizes a Permanent Magnet Synchronous Generator (PMSG) due to its reliability and robustness. Simulation results demonstrate the significant impact of wind speed on rotor speed and electromagnetic torque, highlighting the proportional relationship between wind speed parameters and power output. The controller performance is evaluated using INC and adaptive HCS, with the latter demonstrating superior efficiency under rapid wind speed changes. Additionally, simulation results show that the INC algorithm exhibits rapid tracking capability in approaching the peak maximum power point. Overall, this study provides valuable insights into the performance of MPPT algorithms in WECS, particularly under varying wind conditions.

Keywords: : Incremental conductance algorithm, Hill-climbing search, Wind power, Maximum power point tracking, Permanent magnet synchronous generator, Wind energy conversion system

1. INTRODUCTION

Maximum Power Point Tracking (MPPT) plays a pivotal role in advancing renewable energy systems by maximizing power efficiency within specific operational parameters [1][7]. Beyond technical specifications, its efficacy in reducing installation costs and optimizing power quality during operations is well-documented [8]–[13].

In the realm of photovoltaic (PV) modules, the prevalent approach in literature involves the implementation of either ”hill climbing” or ”incremental conductance” MPPT algorithms due to their simplicity [14], [15]. However, when focusing on innovative MPPT methods, researchers often provide comprehensive summaries of various algorithms [16]–[18]. Critically, research has extensively analyzed the limitations of conventional perturb and observe (P&O) algorithms, particularly regarding perturbation step size selection. To address issues like oscillations and convergence speed, variable step-size P&O algorithms have been developed, classified into modified and adaptive P&O categories.

Figure 1 illustrates a classified model of MPPT with respect to total maximum power captured. Types of MPPT include Direct Power Control (DPC) [11], [19]–[21] and Indirect Power Control (IPC) [1], [2], [8], [11], [22]–[29].

IPC encompasses three approaches: Tip Speed Ratio (TSR), Power Signal Feedback (PSF), and Optimal Torque (OT). The TSR entails an anemometer to measure the speed of the wind as per literature of [30]. The PSF does not require an anemometer but it utilizes turbine blade parameter values.

The OT technique does not require an anemometer either, but still necessitates turbine blade parameter values [11], [28], [31]–[33].

The study by [34] proposes a TSR-based controller to adapt to turbine characteristics near the MPP. Equation (1) defines the relationship between output power (Po) and mechanical power, influenced by generator and converter efficiencies [35], [36]:

$$P_o = \eta_g \eta_c P_{\text{wind}} \qquad (1)$$

This paper focuses on the Hill-Climbing Search (HCS) and Incremental Conductance (INC) algorithms under the DPC controller to enhance power output by operating the turbine closer to its peak maximum point at the DC link [37], [38]. Equation (2) calculates potential power output Pout based

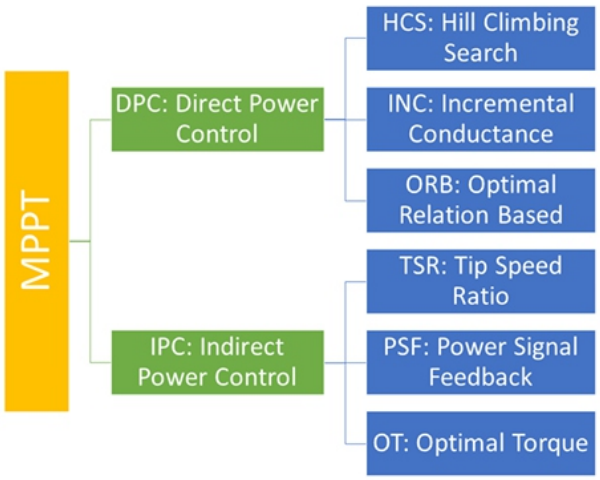


Figure 1. MPPT classified model with respect to total maximum power captured.

on average wind speed [20], [39]:

$$P = 0.5 C_P(\lambda, \beta) \rho \pi R^2 V_w^3, \qquad (2)$$

where ρ is the density of air, Vw is the velocity of wind, R is the radius of rotor, and Cp is the power coefficient [40].

The power coefficient (Cp) depends on blade tip speed (λ) and blade pitch angle (β). As a result of the Betz limit a wind turbine can ideally and theoretically excerpt 59% maximum wind power [39], [41]–[43]. On the other hand, practically the maximum power that can be produced from the wind turbine is up to 40% [44].

Addressing step-size selection, this research introduces a variable step size technique to balance speed control and step-size application efficiently. By monitoring the operating point's distance from the Maximum Power Point (MPP) and transitioning between optimal curves, this technique selects an appropriate step size, enhancing MPP tracking. However, reliance on wind speed measurement and step-size range limitation constrain its effectiveness. The operational methodology leverages adaptive step-size at each operating point, minimizing oscillations and improving WECS performance.
This paper serves as a benchmark for evaluating INC and adaptive HCS algorithms in terms of convergence speed, oscillations, and efficiency performance.

2. DEVELOPMENT OF HILL CLIMBING SEARCH ALGORITHM

A. Conventional Hill Climbing Search Controller

An algorithm such as HCS is used to observe and control the rotors speed variances and output power where an addition to minor decreases or increases to the rotor speed reference. That would eliminate the necessity wind speeds anemometer. Such a controlling technique is adapted through the P&O algorithm. By consistently varying the rotor speed reference, the system will continuously search, leading to a fluctuation in rotor speed referred to as hysteresis around the peak point. The HCS algorithm utilizes electrical power as its input, which is measured using a

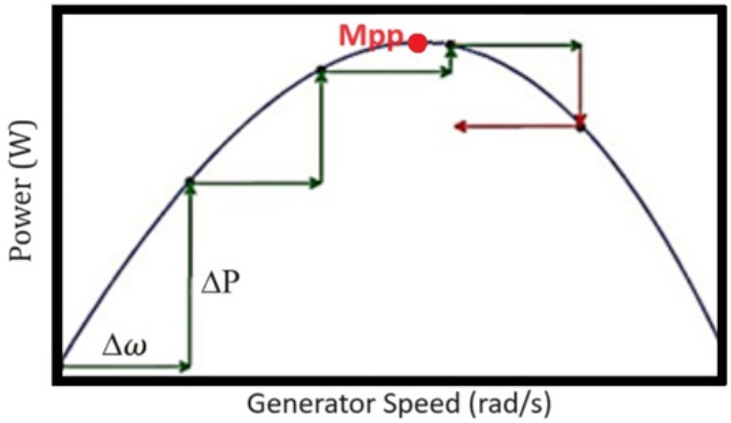


Figure 2. HCS algorithm main principle.

power converter. Realistically, turbine power plays a crucial role during the controlling stage in reaching the peak point. Those powers equally considered when the systems module is at steady-state condition, waiting for a period of time for which the generator powers transient has dissipated [9].

HCS algorithm is a well familiar type of algorithms however it has a very common problem that is weak in reaching MPP of a designated module as well has a slow response time in achieving that. This would result in having a cause of oscillation which leads to losses in power since this algorithm tries to track the MPP and keep trying to reach it without being able to do that and in addition will f luctuate around it [24].

Figure 2 shows HCS principles in tracking the optimal power. Therefore, utilizing the electric power measurement, to estimate the control of the next step. The benefit of over-passing wind turbine data is to guarantee the wind turbine will always operate at its actual MPP, regardless of disparities in the blades external characteristics or other influencing parameters. Though the unique characteristics drive HCS to be the optimal selection in MPPT control at various WECS environments; thus is suitable for wind spee3

conditions that change slowly [7], [45]. Applying large step size perturbations can enhance the convergence speed but at the negative impact of affecting efficiency which is called a trade-off. Raising the convergence speed also results in more oscillations around the maximum power point, because HCS control keeps oscillating around the peak point, leading to inevitable fluctuations. Conversely, a reduced step size can improve proficiency but may cause the controller to slow down and struggle to track the peak point under dynamic changes in wind factors. The determination of the next perturbation stage in the conventional mechanism is based on the power decrease or increase caused by the previous perturbation step. However, this approach can be misleading if the factors of wind change are not taken into account. In such cases, the change in wind can override the effect of the applied perturbation, leading to an incorrect estimation by the traditional algorithm used in HCS. Consequently, the tracking of the peak point becomes insufficient, and the

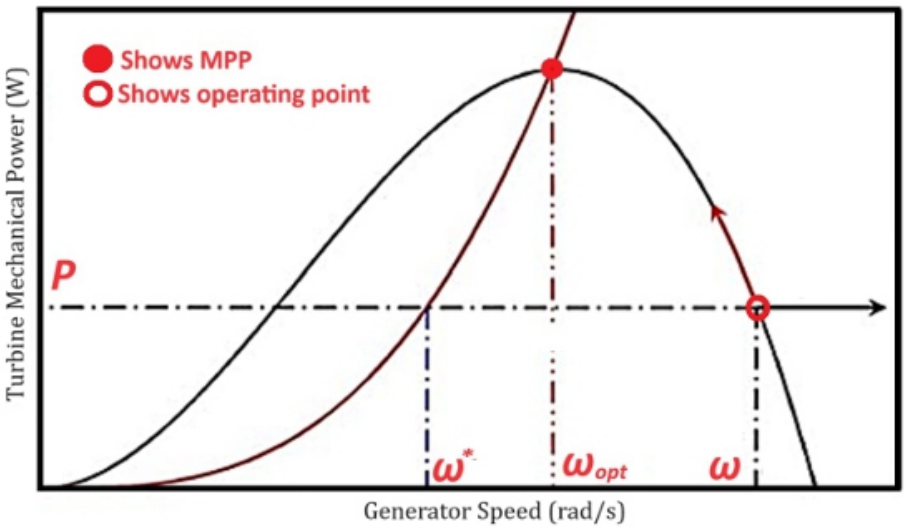


Figure 3. Adaptive HCS.

HCS goes downward. From the literature, researchers have proposed using algorithms with variable-step sizes to get to the optimal peak point. Nevertheless, these algorithms have several drawbacks when the operating point is far from MPP. This is primarily caused by the increased extent of the P−ω slope as shown in Figure 3.

B. Adaptive/HCS-Three Mode

Numerous structures incorporate a primary control unit known as the ”master control” that determines the operational mode of the controller, based on wind speed and disparities in wind speed. Such an approach enables the controller to respond accordingly to minor or significant fluctuations in wind speed, or alternatively, maintain a constant rotor speed within a specified dead band limit.

Figure 3 demonstrates adaptive HCS and shows how to reach the MPP point. The potential for adaptive controllers emerges when they initially operate in a knowledge mode to ascertain the crucial parameters based on a specific wind pattern [7], [46], [47]. Other techniques in the HCS domain are also stated, such as HCS with variable dual step size, as well as search recollect algorithms that hold in a special memory the MPP during the knowledge stage [1], [37], [46]–[52].

C. Hill Climbing Search Adaptive with Power Prediction mode

In the literature Badawi et al 2020 used a novel algorithm that relies on two main mode stages and an intelligent tools which introduced in [9], [10] known as power predicting mode. The introduced algorithm constructed a set of two different modes in detecting the MPP. The main goal of the enhancement is to achieve MPPT in a short time frame. And hence improving the power efficiency of WECS.Note that no iterations calculation is involved and further improving any trade-off as a result of the convergence rate and efficiency [48].

The design and structuring of such algorithm took into consideration how to be simple. It can be applied at different wind speed profiles. Therefore, it can be reached to themost possible power out of the WECS. A novel algorithm can estimate and predict captured power at wind turbine. Thus, the duty cycle can be applied to the MOSFET to detect the optimal point on the real time without delay [8], [29]. Through the "power prediction technique" a division in the range's references for the wind-speed is actioned for the purpose in getting maximum wind-energy. Through this mechanism, speed of wind identifies the (Pout) established at wind-speed range. The results obtained based on the theory would assure that the proposed enhanced algorithm is notable to be fast and further being efficient as compared to a three mode HCS algorithm. In the power prediction there are five main intervals based on the wind speed data ; interval1: less than 2.5m/s which is considered a very low wind speed; interval 2: wind speed range that's initiated at (2.5 m/s) and may reach a value of up to (8 m/s), interval 3: The speed of wind range from an initial value of (8 m/s) to a value that may reach (13.3 m/s); interval 4: Range of the speed of wind may take an initial value of (13.3 m/s) up to an incremental value of up to (20.3 m/s) and finally, interval 5: A dedicated wind speed above 20.3m/s [10].

3. Incremental Conductance algorithm (INC)

INC is a popular algorithm used in tracking the Pmax. Mostly used extensively because of its ease of use and abilities in tacking the MPP. In addition, the INC algorithm is considered to be utilized as baseline and assumed one of the most standardized references as compared with other novel algorithms.

At the initial start of the process, both the voltage (V) and current (I) determine the WECS output. The changes in the values of the above-mentioned parameters (I&V) are identified during the calculation in predicting the I and V derivative. Conventional INC algorithms rely on the basis of comparing the current's derivative value (function of voltage) with the instantaneous current of the WECS against voltage. Simply, the MPP is tracked through this technique by incrementing and decrementing an applied reference voltage in accordance with the current operating point of the module [53]. When this method drives the operating point to reach the MPP and hence to conclude the following [54];

$$\frac{dP}{dV} = 0 \tag{3}$$

$$\frac{dP}{dV} = \frac{d(V \cdot I)}{dV} \cdot dV = I\frac{dV}{dV} + V\frac{dI}{dV} \tag{4}$$

$$\frac{dI}{dV} = -\frac{I}{V}, \tag{5}$$

where, dI/dV is the current derivative, and I/V is the instantaneous PV current to voltage.
Further, this method is based on the slope of P-V slope. The Pmax is reached when it gets a zero slope [55], [56].

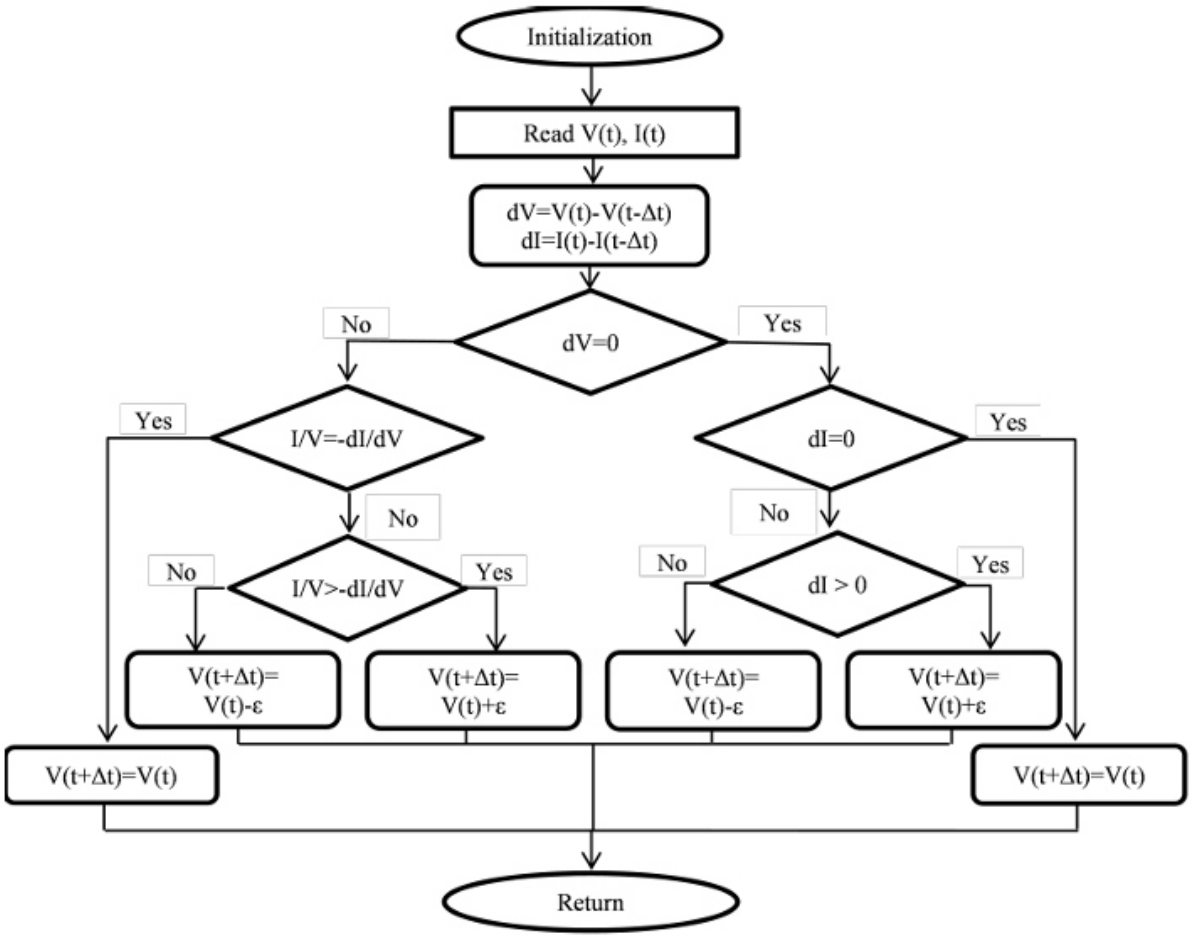


Figure4. INCalgorithmflowchart.

ThebelowequationssumuptheINCscheme[57]:

$$p = VI \tag{6}$$

$$\frac{dp}{dv} = I + V\frac{di}{dv}. \tag{7}$$

Note that when dP dV > 0; I V > −dI dV then in this scenario thevoltagemust be increased (incremented) and viceversawhen dP dV<0; I V<−dI dV thenthevoltagemustbe decreasedor (decremented). INCflowchart is represented in Figure 4. In a conventional context of INC scheme, "ϵ"demonstrates a nominal voltage (fixed) that serves for decrementing and incrementing it accordingly based on the system requirement sand needs.

4. MODELLINGWIND ENERGY CONVERSION SYSTEMBASEDALGORITHMS

The INCalgorithmandHCSadaptivealgorithmwere applied to theWECSusingdifferentwind speedprofiles with rapid changes based on time. In the first part, the (INC) controller isapplied toWECStoobtainPmax.The followingmodel represents themain components in the systemasexposedinFigure5.Thesynchronousgenerator dynamicmodel used in this study is a result of twokey phases' reference synchronous; (d) direct and the other one is (q) quadrature axis frame. Angle formed angle amongthosetwotermsisroughlyaround90ºdegrees.This angle is estimated tobe thedirectionof rotation.Thedq transformationthatwasutilizedinthethree-PhaseofWECS is illustrated throughEquation(8) [58].Note that another illustrationof theinversetransformisreflectedinEquation (9) [1], [8], [26], [40], [59]–[64].

$$\begin{bmatrix} F_d \\ F_q \end{bmatrix} = \frac{2}{3}\begin{bmatrix} \sin\omega t & \sin(\omega t - \frac{2\pi}{3}) & \sin(\omega t + \frac{2\pi}{3}) \\ \cos\omega t & \cos(\omega t - \frac{2\pi}{3}) & \cos(\omega t + \frac{2\pi}{3}) \end{bmatrix}\begin{bmatrix} F_a \\ F_b \\ F_c \end{bmatrix}, \quad (8)$$

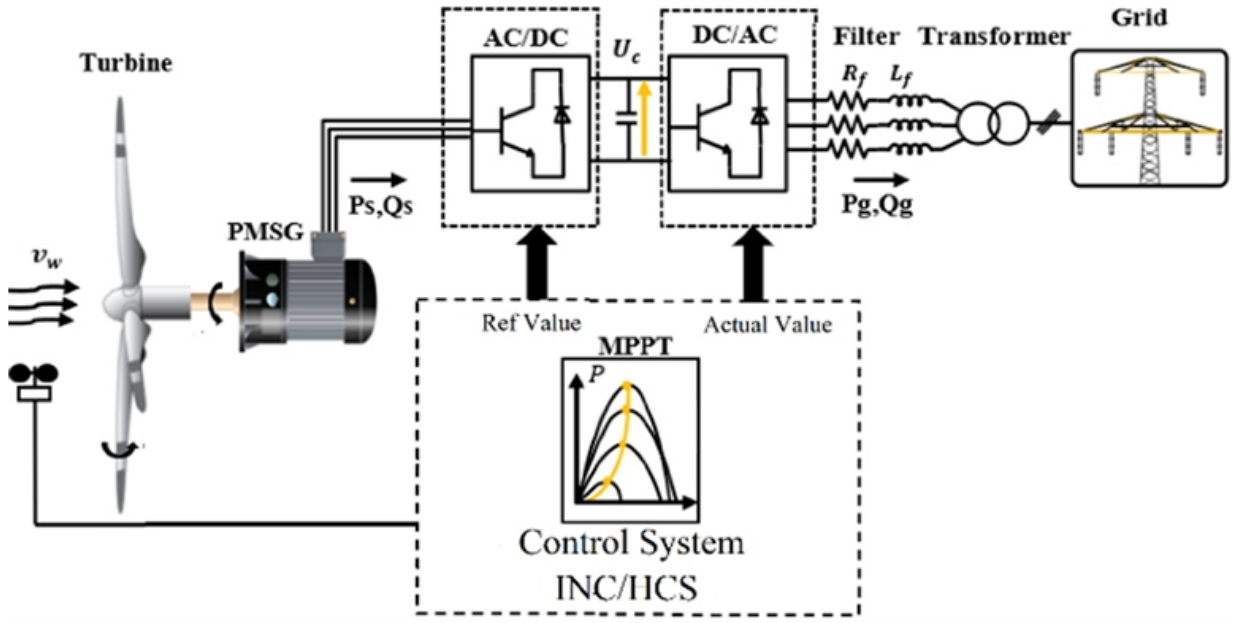


Figure 5. PMSG wind turbine controller algorithm.

$$\begin{bmatrix} F_a \\ F_b \\ F_c \end{bmatrix} = \begin{bmatrix} \sin\omega t & \cos\omega t \\ \sin(\omega t - \frac{2\pi}{3}) & \cos(\omega t - \frac{2\pi}{3}) \\ \sin(\omega t + \frac{2\pi}{3}) & \cos(\omega t + \frac{2\pi}{3}) \end{bmatrix}\begin{bmatrix} F_d \\ F_q \end{bmatrix}. \quad (9)$$

In the frequency domain, the synchronous generator stator current'sd-axis illustrated inEquation(10) and the same applies toq-axis[8].

$$i_{qs} = \frac{\begin{pmatrix} -v_{qs} & -R_s i_{qs} & -\omega_r (L_{ds} + L_{ls}) \end{pmatrix} \begin{pmatrix} i_{ds} \\ \omega_r \phi_r \end{pmatrix}}{S\,(L_{ds} + L_{qs})}, \qquad (10)$$

where, ids is the stator currents d-axis, iqs is the stator current'sqaxis,vds/vqs is thestatorvoltaged−qaxes,ωr is theangular speedgenerator,Rs is thestator resistance, ϕr istherotor'sflux,and Lq Ld istheself-inductanceofstator q−daxes.
Thebelowequations are settingup theparameters of (Ld)and(Lq) respectively.

$$L_d = L_{ls} + L_{dm} \qquad (11)$$

$$L_q = L_{ls} + L_{qm} \qquad (12)$$

Knowing that (Lls) is inductance leakage and (Ldm) and (Lqm) representing in this schemea magnetized inductance at(dq-axes) under the influence of a synchronous generator. (dq)-axis streamlined model in synchronous frame rotor field is addresse din[8].
The calculations of both parameters (ωr) and (Te) of PMSG are reflected at Equations(13)and(14) [8];

$$T_e = \frac{3N_p}{2}\left(\phi_r i_{qs} - (L_d - L_q) i_{ds} i_{qs}\right), \qquad (13)$$

$$\omega_r = \frac{N_p}{JS}(T_e - T_m). \qquad (14)$$

where;Npp is the pole pairs numbers,Tm is the generator mechanical torque, and J is the inertia rotation. Another set of terms to define are the generator speed (pu) and synchronous generator torque Tm(pu). As per the rotor of PMSG, an applied torque can be addressed as follows [8], [65], [66].

$$T_m = \frac{0.5 C_P(\lambda, \beta) \rho \pi R^2 V_w^3}{\omega_r}. \qquad (15)$$

At a certain model, note that (β) is given a determined value however it is assigned a value of β = 0; such a value is used in wind turbine with a trivial scale.

$$\lambda = \frac{V_{\text{tip}}}{V_w} = \frac{\omega_r}{V_w}. \qquad (16)$$

From the derived Equations (2) and (16), wind turbine output maximized power is calculated using Equation (17), and constant optimal wind is calculated through Equations (18) and (19).

$$P_{\max} = K_{\text{opt}}\omega_{\text{ropt}}^{3}, \tag{17}$$

$$K_{\text{opt}} = \frac{0.5\pi\rho C_{\text{pmax}}R^{5}}{\lambda_{\text{opt}}^{3}}, \tag{18}$$

$$\omega_{\text{opt}} = \frac{\lambda_{\text{opt}}V_{\omega}}{R}. \tag{19}$$

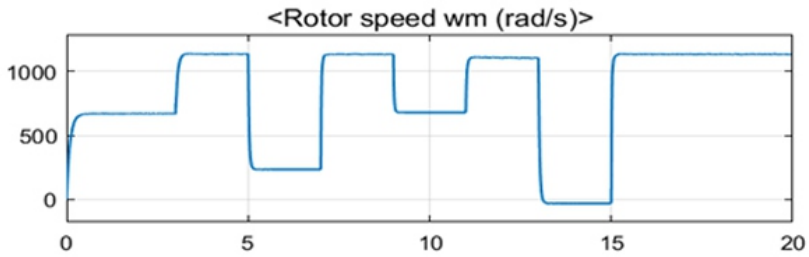
In exceptional cases formulations, Equation (19) is used to determine (ωopt) when a condition is applied where the rated speed at both the PMSG and MPPT [8], [24]. Reaching a unity power is a challenging task when it comes to micro controllers where adjusting the module parameters may have a great effect on the overall performance.

Again, when rated speed was addressed, an assurance is required to ensure this is ideally reaching PMSG along with the power factor (unity stage) and a dc-dc boost converter can be controlled at this case by Duty cycle. The overall advantage is to collect the maximum power available from WECS [67]–[70].

MPPT algorithm drives the integrated module of wind turbine to the highest and ultimate possible speed denoted and addressed as ωopt for every and single wind potential velocity. As a result of the preceding discussions, we may conclude at this stage that, arriving towards the maximum power point will solely be dependable purely on the MPPT scheme control [71], [72].

5. RESULTS AND DISCUSSION

PMSG's induction generator type with WECS model connected with the controller (INC/ HCS) Adaptive to check the performance effectiveness in reaching MPP. The 3-phase output voltage from PMSG is rectified in the sense of converting an AC power waveform to its DC component and then feed it into a predefined controller. To control the voltage (Vdc) a dc-dc converter unit is designated for this task. Where a reference voltage (Vref) is supplied by the MPPT controller. This supplied (Vref) is driven to be compared with the (Vdc) value and eventually, the final value's reading goes into the controller. Knowing that at this



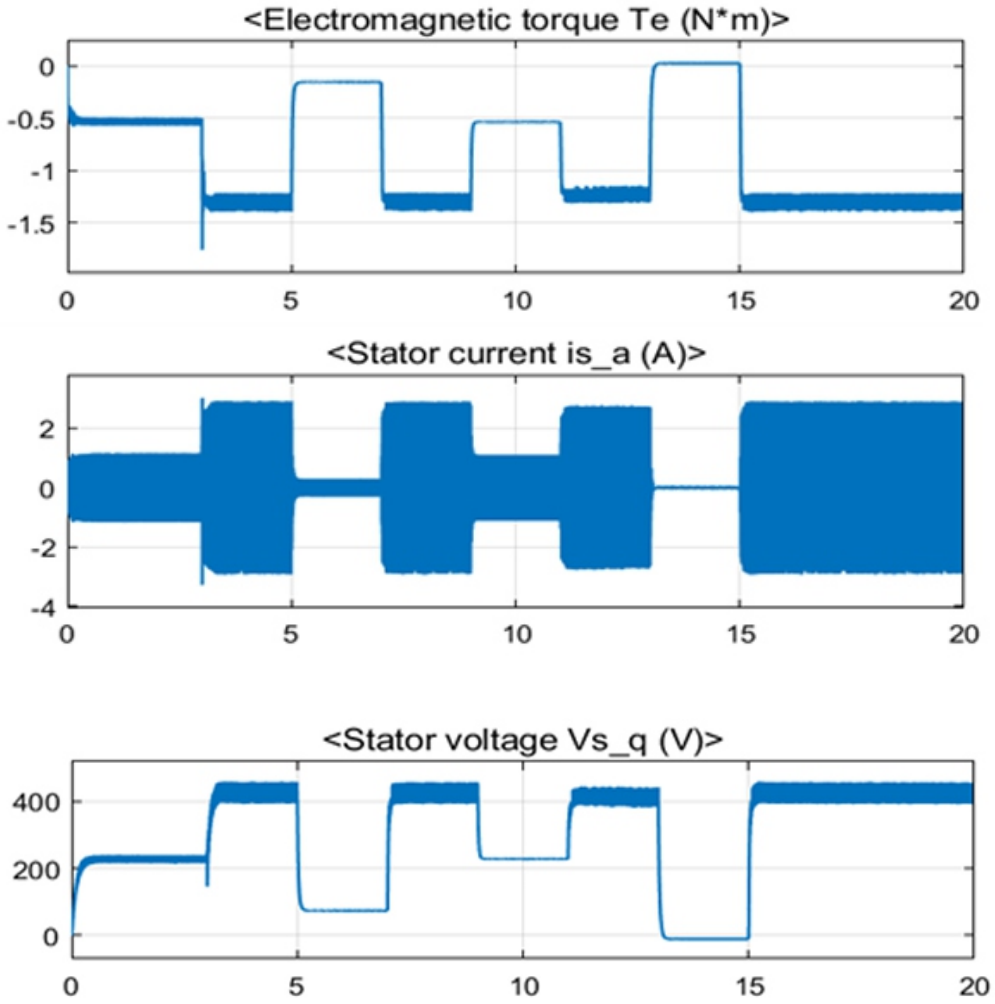


Figure 6. (Stators I&V), speed of rotor, and WECS electromag netic/torque.

point the controllers' output is checked against the (k − 1) state in a forward step to seek the best switching process of a DC-DC convertor, i.e. the "ON"/"OFF" states. From Figure 6 the rotor speed increased rapidly because of wind speeds' raised value. The electromagnetic torque is following the rotor speed based on wind speed value.

Stator voltage and current increased dramatically by mean of increasing the mean wind speed data and as mentioned in the literature [10], [24], [29], [56] the mean wind speed has cubic proportional relation with the power value. It can be noticed that the rotor speed value changed due to the wind speed data. The electromagnetic torque has inverse proportional relation with the rotor speed. This is displayed in Figure 6. Stator (I) and (V) raised intensely as per the data of winds' speed and followed the rotor speed behaviour which is due to the proportional relations.

Figure 7 shows a sinusoidal 3-phase power output for the PMSG. It is noticed that the amplitude was increasing due to the fact that the speed of the mean wind is incrementing. The three-phase power value followed the (I) and (V) in the phase angle value, where the phase angle is stable without notable leading/lagging issues. Power factor is symbolized by PF = cos(ϕ) = 1 where, in this case the rated power value can be captured from the wind turbine. Therefore,

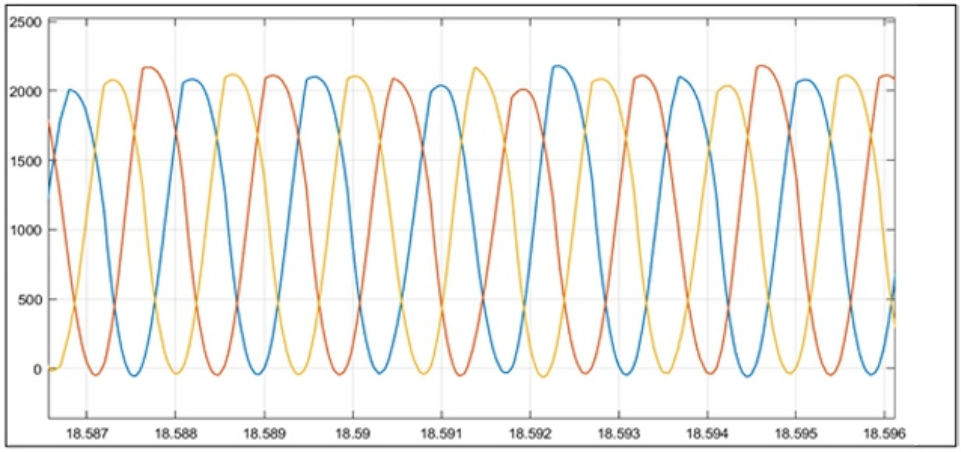


Figure 7. Three-phase power value of the induction generator Pabc between 18.587s and 18.596s power generator.

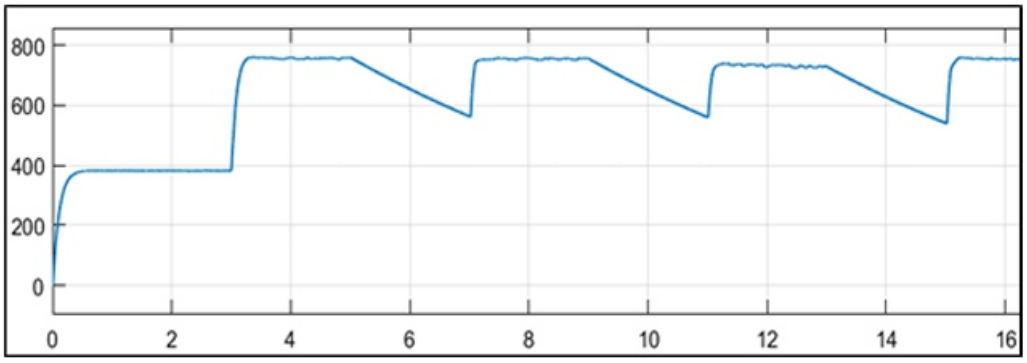


Figure 8. DC voltage illustration at the incident of 3-phase bridge rectifier (no controller involved).

the system in this situation is considered ready to rectify the power output to utilize the controller on the DC linkage that belongs to the boost DC-DC converter. Figure 8 represents a rectified signal from the PMSG using the 3-diode rectifier bridge (no controller), where DC voltage value decreased steadily because of decrease wind speed. In addition, the DC voltage raised sharply when the wind speed increases because of cubic proportional relation taking place through wind speed and power as shown in Equation (1). It can be noticed from the Figure 8 the f luctuation was due to the instability of the applied wind speed. Two different controllers will be applied to this signal to assess how the performance of efficiency acts as the next f igures exhibit.

In Figure 9 INC controller had been applied to the rectified signal to reach the optimal peak point and to enhance the efficiency performance. The result shows fast-tracking capability in detecting the optimal operating point on the power curve, as compared to Figure 8 which shows fluctuations in the captured power.

However, the efficiency performance decreased dramatically based on rapid changes in wind speed. Based on Figure 9, the INC technique has a minimal ripple observed at 7.2 sec, and at 15.3 sec. It can be seen the curve raised roughly at 3.1 sec due to the raised wind speed from 6 m/s to 13 m/s. Here, the incremental technique provided lower oscillation as compared with adaptive HCS.

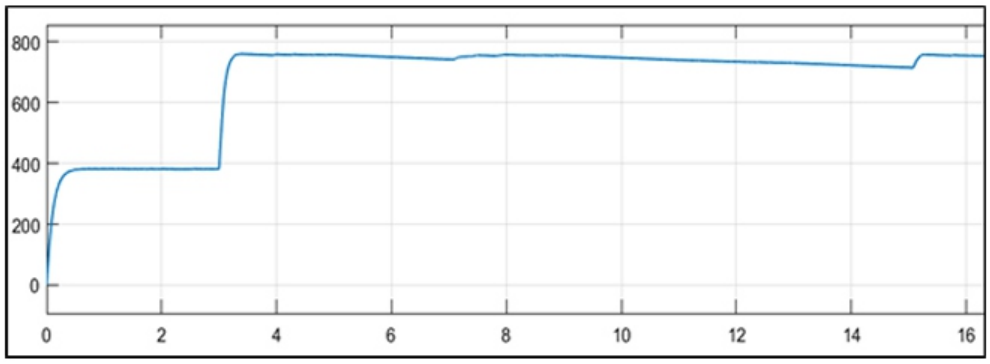


Figure 9. DC voltage after the INC controller.

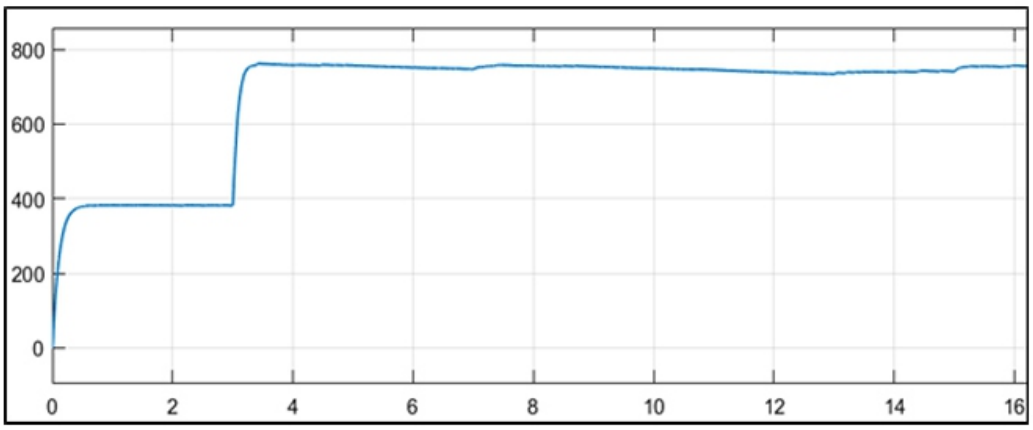


Figure 10. DC voltage after HCS Adaptive controller.

similar efficiency performance as shown in Figures 8 and 9. However, the adaptive HCS has higher captured power compared to the INC, particularly during the rapid change in wind speed.

In Figure 10, it shows the DC volt signal after applying the adaptive HCS. It can be noticed that the adaptive HCS algorithm with power prediction mode along a rapid response upheld the Pmax value successfully which offers a valuable insight into power efficiency. Adaptive HCS algorithm shows the best efficiency performance under rapid change wind speed. Adaptive HCS at every wind speed value, the operating condition is kept at its optimal point. This is due to the power prediction mode stage with high accuracy as compared to incremental technique.

Thus adaptive HCS technique captures Pmax even during wind speeds' dynamic fluctuations, as shown in Figure 10. Further, adaptive HCS technique has a lower overshoot than the incremental technique from 6-8 seconds and 14-16 seconds through wind speed instabilities.

6. Conclusion

In this study, we have addressed two essential algorithms within control structured techniques for wind turbines, integral to tuning WECS with the unique goal of driving PMSG to optimal efficiency by achieving unity power factor. The INC and adaptive HCS algorithms have demonstrated accurate tracking and detection of the MPP. Theoretical results confirm that INC exhibits fast tracking capability

to reach the MPP, albeit with decreased efficiency performance under rapid changes in wind speed. Conversely, adaptive HCS demonstrates higher efficiency and performance in response to rapid changes in wind speed, with enhancements observed in steady-state efficiency.

These findings underscore the importance of algorithm selection and adaptation in optimizing WECS performance under varying wind conditions, contributing to the advancement of renewable energy systems. Further research could explore refinement and integration of these algorithms to enhance overall efficiency and stability in wind energy applications.

References

[1] M. Monica, P. Sivakumar, S. J. Isac, and K. Ranjitha, "Pmsg based wecs: Control techniques, mppt methods and control strategies for standalone battery integrated system," Eigth International Conference on Nin the Applications of Differential Equations in Sciences (NTADES2021), 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:247982668

[2] C. V. Govinda, S. V. Udhay, C. Rani, Y. Wang, and K. Busawon, "A review on various MPPT techniques for wind energy conversion system," 2018 Internat2018 International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC)ional conference on computation of power, energy, Information and Communication (ICCPEIC), pp. 310–326, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:53280766

[3] M. Premkumar, R. Sowmya, C. Ramakrishnan, P. Jangir, E. H. Houssein, S. Deb, and N. Manoj Kumar, "An efficient and reliable scheduling algorithm for unit commitment scheme in microgrid systems using enhanced mixed integer particle swarm optimizer considering uncertainties," Energy Reports, vol. 9, pp. 1029–1053, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352484722026245

[4] J. Hussain and M. K. Mishra, "Adaptive maximum power point tracking control algorithm for wind energy conversion systems," IEEE Transactions on Energy Conversion, vol. 31, no. 2, pp. 697705, 2016.

[5] C. Huang, F. Li, and Z. Jin, "Maximum power point tracking strategy for large-scale wind generation systems considering wind turbine dynamics," IEEE Transactions on Industrial Electronics, vol. 62, no. 4, pp. 2530–2539, 2015.

[6] M. Narayana, G. A. Putrus, M. G. Jovanovi´c, P. S. Leung, and S. P. McDonald, "Generic maximum power point tracking controller for small-scale wind turbines," Renewable Energy, vol. 44, pp. 72–79, 2012. [Online]. Available: https://api.semanticscholar.org/CorpusID: 15207256

[7] S. M. R. Kazmi, H. Goto, H.-J. Guo, and O. Ichinokura, "A novel algorithm for fast and efficient speed-sensorless maximum power point tracking in wind energy conversion systems," IEEE Transactions on Industrial Electronics, vol. 58, no. 1, pp. 29–36, 2011.

[8] A. Badawi, "Maximum power point tracking control scheme for small scale wind turbine," Ph.D. dissertation, International Islamic University Malaysia (IIUM), 2019.

[9] A. Badawi, N. F. Hasbullah, S. H. Yusoff, A. H. A. Hashim, and A. M. Zyoud, "Novel technique for hill climbing search to reach maximum power point tracking," International Journal of Power Electronics and Drive Systems, vol. 11, pp. 2019–2029, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID: 225005618

[10] A. Badawi, N. F. Hasbullah, S. H. Yusoff, A. H. A. Hashim, S. Khan, and A. M. Zyoud, "Power prediction mode technique for hill climbing search algorithm to reach the maximum power point tracking," 2020 2nd International Conference on Electrical, Control and Instrumentation Engineering (ICECIE), pp. 1–7, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:230997525

[11] D. Kumar and K. Chatterjee, "A review of conventional and advanced MPPT algorithms for wind

energy systems," Renewable and Sustainable Energy Reviews, vol. 55, pp. 957–970, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032115012654

[12] E. H. Dursun and A. A. Kulaksiz, "Second-order sliding mode voltage-regulator for improving MPPT efficiency of PMSG-based WECS," International Journal of Electrical Power & Energy Systems, vol. 121, p. 106149, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0142061519339481

[13] M. Premkumar, N. Shankar, R. Sowmya, P. Jangir, C. Kumar, L. M. Abualigah, and B. Derebew, "A reliable optimization framework for parameter identification of single-diode solar photovoltaic model using weighted velocity-guided grey wolf optimization algorithm and Lambert-W function," IET Renewable Power Generation, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:259633486

[14] M. Miyatake, M. Veerachary, F. Toriumi, N. Fujii, and H. Ko, "Maximum power point tracking of multiple photovoltaic arrays: A PSO approach," IEEE Transactions on Aerospace and Electronic Systems, vol. 47, no. 1, pp. 367–380, 2011.

[15] L. Cristaldi, M. Faifer, M. Rossi, and S. Toscani, "An improved model-based maximum power point tracker for photovoltaic panels," IEEE Transactions on Instrumentation and Measurement, vol. 63, pp. 63–71, 2014. [Online]. Available: https://api.semanticscholar. org/CorpusID:35695079

[16] H.Li, D. Yang, W. Su, J. L¨u, and X. Yu, "An overall distribution particle swarm optimization MPPT algorithm for photovoltaic system under partial shading," IEEE Transactions on Industrial Electronics, vol. 66, no. 1, pp. 265–275, 2019.

[17] A. B. M. Elzein, H. Ali, and K. M. A. M. Zyoud, "State of the art perturb and observe MPPT algorithms based wind energy conversion systems: A technology review," Unpublished, 2024.

[18] M. Dhimish, "Assessing MPPT techniques on hot-spotted and partially shaded photovoltaic modules: Comprehensive review based on experimental data," IEEE Transactions on Electron Devices, vol. 66, no. 3, pp. 1132–1144, 2019.

[19] Y. Errami, M. Benchagra, M. Hilal, M. Maaroufi, and M. Ouassaid, "Control strategy for PMSG wind farm based on MPPT and direct power control," 2012 International Conference on Multimedia Computing and Systems, pp. 1125–1130, 2012. [Online]. Available: https://api.semanticscholar.org/CorpusID:16306486

[20] J. Singh and M. Ouhrouche, MPPT control methods in wind energy conversion systems. InTech, Jun. 2011.

[21] M. Kermadi, S. Mekhilef, Z. Salam, J. Ahmed, and E. M. Berkouk, "Assessment of maximum power point trackers performance using direct and indirect control methods," International Transactions on Electrical Energy Systems, vol. 30, 08 2020.

[22] R. Tiwari, K. Kumar, N. R. Babu, and K. R. Prabhu, "Coordinated mppt and dpc strategies for PMSG based grid connected wind energy conversion system," Energy Procedia, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:115279568

[23] M. B. Toriki, M. K. Asy'ari, and A. Musyafa, "Enhanced performance of PMSG in WECS using MPPT- Fuzzy sliding mode control," Journal Europ´ een des Syst`emes Automatis´es, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID: 233782872

[24] A. Badawi, N. F. Hasbullah, S. H. Yusoff, A. H. A. Hashim, S. Khan, and A. M. Zyoud, "Paper review: maximum power point tracking for wind energy conversion system," 2020 2nd International Conference on Electrical, Control and Instrumentation Engineering (ICECIE), pp. 1–6, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:230994725

[25] A. Badawi, H. Ali, N. A. Ismail, P. Ramallah, A. Zyoud, and S. H. Yusoff, "Weibull probability distribution based on four years wind speed data using nine numerical methods," Unpublished.

[26] J. Baran and A. Jaderko, "An mppt control of a PMSG-based WECS with disturbance compensation and wind speed estimation," Energies, 2020. [Online]. Available: https://api.semanticscholar.org/ CorpusID:229431810

[27] H. H. H. Mousa, A. Youssef, and E. E. M. Mohamed, "State of the art perturb and observe MPPT algorithms based wind energy conversion systems: A technology review," International Journal of Electrical Power & Energy Systems, vol. 126, p. 106598, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID: 228907082

[28] R. Maher, A. K. Abdelsalam, Y. G. Dessouky, and A. Nouman, "High performance state-flow based mppt technique for micro WECS," IET Renewable Power Generation, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:208832514

[29] A. Badawi, H. Ali, N. A. Ismail, P. Ramallah, A. Zyoud, and S. H. Yusoff, "Wind energy production using novel HCS algorithm to reach MPPT for small-scale wind turbines under rapid change wind speed," Unpublished.

[30] G. Hua and Y. Geng, "A novel control strategy of MPPT taking dynamics of wind turbine into account," in 2006 37th IEEE Power Electronics Specialists Conference, 2006, pp. 1–6.

[31] K. S. M. Raza, H. Goto, H.-J. Guo, and O. Ichinokura, "Maximum power point tracking control and voltage regulation of a dc gridtied wind energy conversion system based on a novel permanent magnet reluctance generator," in 2007 International Conference on Electrical Machines and Systems (ICEMS), 2007, pp. 1533–1538.

[32] I. Buehring and L. Freris, "Control policies for wind-energy conversion systems," Generation, Transmission and Distribution, IEE Proceedings C, vol. 128, pp. 253– 261, 10 1981.

[33] S. M. R. Kazmi, H. Goto, H.-J. Guo, and O. Ichinokura, "Review and critical analysis of the research papers published till date on maximum power point tracking in wind energy conversion system," in 2010 IEEE Energy Conversion Congress and Exposition, 2010, pp. 4075–4082.

[34] A. Mahdi, W. Tang, L. Jiang, and Q. Wu, "A comparative study on variable-speed operations of a wind generation system using vector control," Renewable Energy and Power Quality Journal, vol. 1, 04 2010.

[35] A. Badawi, H. Ali, I. M. Elzein, A. M. Zyoud, and A. AbuHudrouss, "Highly efficient pure sine wave inverter using microcontroller for photovoltaic applications," 2023 International Symposium on Networks, Computers and Communications (ISNCC), pp. 1–6, 2023. [Online]. Available: https://api.semanticscholar.org/ CorpusID:265484054

[36] S. Agrawal, S. Pandya, P. Jangir, K. Kalita, and S. Chakraborty, "A multi-objective thermal exchange optimization model for solving optimal power flow problems in hybrid power systems," Decision Analytics Journal, vol. 8, p. 100299, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/ pii/S277266222300139X

[37] H. Gouabi, A. Hazzab, M. Habbab, M. Rezkallah, and A. Chandra, "Experimental implementation of a novel scheduling algorithm for adaptive and modified po mppt controller using fuzzy logic for WECS," International Journal of Adaptive Control and Signal Processing, vol. 35, 06 2021.

[38] L. Mdakane and M. Kamper, "Simple robust MPPT control for wind energy DC-grid connected systems," in IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society, Oct 2019, pp. 2495–2500.

[39] A. Badawi, "An analytical study for establishment of wind farms in Palestine to reach the optimum electrical energy," 07 2013.

[40] A. Badawi, N. F. Hasbullaha, S. H. Yusoff, S. Khan, A. H. A. Hashim, A. M. Zyoud, and M. Elamassie, "Evaluation of wind power for electrical energy generation in the mediterranean coast of Palestine for 14 years," International Journal of Electrical and Computer Engineering (IJECE), 2019.

[Online]. Available: https://api.semanticscholar.org/CorpusID:208083439

[41] A. Badawi, S. Yusoff, A. Zyoud, S. Khan, A. Hashim, Y. Uyaro˘glu, and M. Ismail, "Data bank: nine numerical methods for determining the parameters of weibull for wind energy generation tested by five statistical tools," International Journal of Power Electronics and Drive Systems (IJPEDS), vol. 12, p. 1114, 06 2021.

[42] A. Badawi, N. F. Hasbullah, S. H. Yusoff, A. H. A. Hashim, and M. Elamassie, "Practical electrical energy production to solve the shortage in electricity in Palestine and pay back period," International Journal of Electrical and Computer Engineering, vol. 9, pp. 4610–4616, 2019. [Online]. Available: https://api. semanticscholar.org/CorpusID:208076671

[43] A. Badawi, M. Ouda, A. M. Zyoud, and S. H. Yusoff, "The simplest estimation method of weibull probability distribution parameters," 2021 6th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE), vol. 6, pp. 1–5, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID: 246870607

[44] L. P. Chamorro and R. E. A. Arndt, "Non-uniform velocity distribution effect on the Betz–Joukowsky limit," Wind Energy, vol. 16, pp. 279–282, 2013. [Online]. Available: https://api. semanticscholar.org/CorpusID:120823911

[45] S. Kumar, P. Jangir, G. G. Tejani, and M. Premkumar, "A decomposition based multi-objective heat transfer search algorithm for structure optimization," Knowledge-Based Systems, vol. 253, p. 109591, 2022. [Online]. Available: https://www.sciencedirect.com/ science/article/pii/S0950705122008024

[46] R. D. Shukla and R. K. Tripathi, "Maximum power extraction schemes & power control in wind energy conversion system, International Journal of Scientific and Engineering Research, vol. 3, 06 2012.

[47] J. Singh and M. Ouhrouche, MPPT control methods in wind energy conversion systems. InTech, Jun. 2011.

[48] H. Mousa, A.-R. Youssef, and E. Mohamed, "Study of robust adaptive step-sizes P&O MPPT algorithm for high-inertia wt with directdriven multiphase pmsg," International Transactions on Electrical Energy Systems, 07 2019.

[49] M. C. Akkaya, A. Polat, and L. T. Ergene, "Mppt based adaptive control algorithm for small scale wind energy conversion systems with pmsg," in 2019 International Aegean Conference on Electrical Machines and Power Electronics (ACEMP) 2019 International Conference on Optimization of Electrical and Electronic Equipment (OPTIM), 2019, pp. 517–522.

[50] Z. Alrowaili, M. Ali, A.-R. Youssef, H. Mousa, A. Ali, G. AbdelJaber, M. Ezzeldien, and F. Gami, "Robust adaptive HCS MPPT algorithm-based wind generation system using model reference adaptive control," Sensors, vol. 21, 07 2021.

[51] A. Yesudhas, Y. H. Joo, and S. Lee, "Reference model adaptive control scheme on PMVG-based WECS for MPPT under a real wind speed," Energies, vol. 15, p. 3091, 04 2022.

[52] M. M. Ali, A. Youssef, A. S. Ali, and G. T. Abdel-Jaber, "Variable step size PO MPPT algorithm using model reference adaptive control for optimal power extraction," International Transactions on Electrical Energy Systems, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:201235921

[53] R. I. Putri, S. Wibowo, and M. Rifa'i, "Maximum power point tracking for photovoltaic using incremental conductance method," Energy Procedia, vol. 68, pp. 22–30, 2015, 2nd International Conference on Sustainable Energy Engineering and Application (ICSEEA) 2014 Sustainable Energy for Green Mobility. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1876610215005342

[54] M. H. Ibrahim, S. P. Ang, M. N. Dani, M. I. Rahman, R. Petra, and S. M. Sulthan, "Optimizing step-

size of perturb & observe and incremental conductance MPPT techniques using PSO for grid-tied pv system," IEEE Access, vol. 11, pp. 13079–13090, 2023.

[55] M. N. Ali, K. Mahmoud, M. Lehtonen, and M. M. F. Darwish, "An efficient fuzzy-logic based variable-step incremental conductance MPPTmethod for grid-connected PV systems," IEEE Access, vol. 9, pp. 26420–26430, 2021.

[56] A. Badawi, M. Ouda, A. M. Zyoud, and S. H. Yusoff, "Maximum power point tracking controller technique using permanent magnet synchronous generator," 2021 6th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE), vol. 6, pp. 1–5, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:246870229

[57] T. Esram and P. Chapman, "Comparison of photovoltaic array maximum power point tracking techniques," Energy Conversion, IEEE Transactions on, vol. 22, pp. 439– 449, 07 2007.

[58] P. Krause, O. Wasynczuk, S. D. Sudhoff, and S. Pekarek, Analysis of electric machinery and drive systems, 3rd ed., ser. IEEE Press Series on Power Engineering, P. Krause, O. Wasynczuk, S. Sudhoff, and S. Pekarek, Eds. Nashville, TN: John Wiley & Sons, Aug. 2013.

[59] K. Amei, Y. Takayasu, T. Ohji, and M. Sakui, "A maximum power control of wind generator system using a permanent magnet synchronous generator and a boost chopper circuit," Proceedings of the Power Conversion Conference-Osaka 2002 (Cat. No.02TH8579), vol. 3, pp. 1447–1452 vol.3, 2002. [Online]. Available: https://api.semanticscholar.org/CorpusID:110635579

[60] A. Jain, S. Shankar, and V. Vanitha, "Power generation using permanent magnet synchronous generator (PMSG) based variable speed wind energy conversion system (WECS): An overview," Journal of green engineering, vol. 7, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:108402216

[61] A. Westlake, J. R. Bumby, and E. S. Spooner, "Damping the power-angle oscillations of a permanent-magnet synchronous generator with particular reference to wind turbine applications," 1996.

[62] A. Badawi, N. Hasbullah, S. Yusoff, S. Khan, A. Hashim, A. Zyoud, and M. Elamassie, "Weibull probability distribution of wind speed for Gaza strip for 10 years," Applied Mechanics and Materials, vol. 892, pp. 284–291, 06 2019.

[63] A. Badawi, N. Hasbullah, S. Yusoff, and A. Hashim, "Energy and power estimation for three different locations in Palestine," Indonesian Journal of Electrical Engineering and Computer Science, vol. 14, pp. 1049–1056, 06 2019. [64] N. Verma, S. Banerjee, S. Gupta, S. Goyal, and R. Sharma, "PMSG based WECS with MPPT via modified p & o algorithm," in 2019 3rd International Conference on Recent Developments in Control, Automation & Power Engineering (RDCAPE), 2019, pp. 646–650.

[65] P. Sahin, R. Resmi, and V. Vanitha, "Pmsg based standalone wind electric conversion system with mppt," in 2016 International Conference on Emerging Technological Trends (ICETT), 2016, pp. 1–5.

[66] U. H. Khan, Q. Khan, L. Khan, W. Alam, N. Ali, I. Khan, K. S. Nisar, and R. A. Khan, "Mppt control paradigms for PMSG-WECS: A synergistic control strategy with gain-scheduled sliding mode observer," IEEE Access, vol. 9, pp. 139876–139887, 2021.

[67] Y. Errami, M. Maaroufi, and M. Ouassaid, "A MPPT vector control of electric network connected wind energy conversion system employing PM synchronous generator," 2013 International Renewable and Sustainable Energy Conference (IRSEC), pp. 228–233, 2013. [Online]. Available: https://api.semanticscholar.org/ CorpusID:8330996

[68] Y. Errami, M. Hilal, M. Benchagra, M. Maaroufi, and M. Ouassaid, "Nonlinear control of MPPT and grid connected for wind power generation systems based on the PMSG," 2012 International Conference on Multimedia Computing and Systems, pp. 1055–1060, 2012. [Online]. Available: https://api.semanticscholar.org/CorpusID: 8774171

[69] E. Radwan, S. Kamel, L. S. Nasrat, and A. Youssef, "Improved PO MPPT for grid connected wind energy conversion system," 2023 IEEE Conference on Power Electronics and Renewable Energy (CPERE), pp. 1–6, 2023. [Online]. Available: https: //api.semanticscholar.org/CorpusID:258641711

[70] A. Badawi, S. Kazmi, R. Boby, M. Shah, and K. Matter, "Resonant circuit response for contactless energy transfer under variable pwm," International Journal of Information and Electronics Engineering, vol. 7, pp. 41–47, 01 2017.

[71] J. Chen, W. Yao, Q. Lu, Y. Ren, W. Duan, J. Kan, and L. Jiang, "Adaptive active fault-tolerant MPPT control of variable-speed wind turbine considering generator actuator failure," International Journal of Electrical Power Energy Systems, vol. 143, p. 108443, 2022. [Online]. Available: https: //www.sciencedirect.com/science/article/pii/S0142061522004537

[72] M. Beghdadi and K. Kouzi, "Novel fully sensorless synergetic control of brushless doubly fed induction machine integrated in wind energy conversion system driven by fuzzy-based HCS MPPT algorithm," Wind Engineering, 10 2022.

Ahmed Badawi is an Assistant Professor at the University of Doha for Science and Technology (UDST), Qatar. Dr. Badawi holds a PhD in Electrical Engineering from the International Islamic University Malaysia (IIUM). Dr. Badawi has been honored with several awards and scholarships throughout his academic career. He has contributed to numerous high-ranking journals in Electrical Engineering and renewable energy with his published papers. Dr. Badawi possesses extensive expertise in renewable energy, power control systems, as well as automation and control systems.He serves as a Reviewer for Energy Elsevier and IET Power Generation journals.

Hassan Ali holds a PhD in Electrical Engineering from the University of Melbourne, Australia, and currently serves as an Assistant Professor (Electrical Engineering) at UDST, Qatar. With extensive experience in machine learning, Internet of Things, embedded systems, wireless communications, and electrical engineering domains, Dr. Ali is a Senior Member IEEE, USA. He has authored numerous research articles and spearheaded several RD projects focused on pioneering technologies and innovative solutions.

I. M. Elzein is an electrical and electronics engineer who pursued his undergraduate and graduate level degrees in Electrical and com puter engineering from Wayne State Univer sity, Michigan, USA in 2004. Currently, Dr. Elzein is holding an academic lecturing role in the department of telecommunication and networking engineering at the University of Doha for Science and Technology. Dr. Elzein has more than 70 research papers in multi international conferences and journals as well being a reviewer committee member and technical program committee. His research profile is mainly in the fields of electrical and electronic engineer ing with a concentration on (PV) photovoltaic systems, robotics design, renewable energy and information technology.

Alhareth Zyoud received his bachelor's de gree in Electrical Engineering from Palestine Polytechnic University in 2006, followed by his master's and Ph.D. degrees in Com munication Engineering from the Interna tional Islamic University Malaysia in 2011 and 2017, respectively. He has authored or co-authored numerous research papers pub lished in international journals and confer ences. His current research interests include RF modeling and simulation, 5G radio resource management, and rain attenuation analysis.

ISSN 2210-142X

A Survey on the MT Methods for Indian Languages: MT Challenges, Availability, and Production of Parallel Corpora, Government Policies and Research Directions

Sudeshna Sani1, Samudra Vijaya2 and Suryakanth V Gangashetty3
1,2,3Department of CSE, Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh, India

ABSTRACT

Since 1991, machine translation has been a prominent research area in India, with IIT Kanpur pioneering the original work which has since been expanded to several universities. Only 10 percent of India's 1.3 billion inhabitants can read, write, and speak English with varying degrees of competence, which makes machine translation crucial in overcoming the linguistic barrier to the internet. The Indian market for commercial products and events is greatly influenced by local languages, making the development and translation of region-based content an essential research topic nowadays. However, Indic-to-Indic language direct translation has faced several challenges and is still going through the experimental phase. Several government-sponsored projects are being undertaken in this regard. Still, there are limited sentence-aligned parallel bi-text resources available for the majority of Indian language pairs. This paper presents a detailed survey of the current trends of research on machine translation between Indian languages, along with their challenges over time. It also presents a timeline of recent research conducted and key findings of past surveys conducted over a decade. Under a single canopy, this paper provides sources of data, the progress made in developing datasets for low-resource Indian languages, various models of translation, encouragement from Indian Govt., and finally, new research directions.

Keywords: Machine Translation, RBMT, SMT, NMT, Low-Resource Indian languages,BLUE, METEOR, AI4Bharat, Bhashini

1. Introduction

Machine Translation (MT) is a method of translating one written human language automatically in to another language, while maintaining the significance of the source text and generating fluent and proper text in the target language. MT has been developed as a subfield of Artificial Intelligence (AI) and is a part of computational linguistics and language engineering. MT techniques are further improved by utilizing concepts and methods from various fields such as statistics, computer science, AI, translation theory, and linguistics [1]. Figure 1 shows the basic structure of an MT system.

Machine Translation (MT) research in Indian languages is relatively less developed as compared to other international languages such as English, Chinese, and Spanish. This is primarily due to the complexity and diversity of Indian languages, which makes MT a challenging task. Additionally, Indian languages have low resource availability, lack of parallel corpora, and limited research funding. However, in recent years, a growing MT research interest

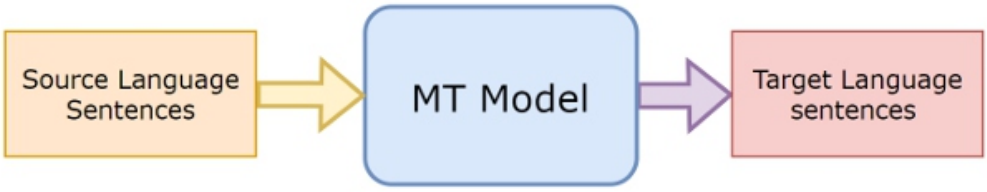


Figure 1. Diagram of a Basic MT System

for Indian languages is observed, with several initiatives and collaborations between academia, industry, and government. Various research projects are underway to advance MT systems for Indian languages, and efforts are being made to increase the availability and quality of parallel corpora for Indian languages. Despite the challenges, MT research in Indian languages has great potential in the current global market scenario. India is a distinct country with more than 1.3 billion residents, and a growing economy with a huge demand for localization of content in regional languages. Indian languages are typically classified into five major language families [2] [3]:

• Indo-European: This family includes languages such as Bengali, Hindi, Gujarati, Marathi, Punjabi and Urdu.
• Dravidian: This family includes languages such as Tamil, Telugu, Kannada and Malayalam.
• Austroasiatic: This family includes languages such as Santali, Khasi and Mundari.
• Sino-Tibetan: Exemplar languages of this family are Manipuri, Lepcha and Bhutia.
• Andamanese: This family includes the languages spoken by the indigenous tribes of the Andaman and Nicobar Islands.

Each of these language families is further divided into numerous subgroups and dialects, reflecting the linguistic diversity of India. India boasts a large diverse linguistic area with more than 22 official languages and over 1,600 mother-tongues [2]. However, only a small percentage of the Indian inhabitants can read, write, and speak English fluently. In the current global market scenario, where businesses and consumers operate on a global scale, language barriers can become a major obstacle for companies trying to reach out to new markets. Machine Translation (MT) technology can help bridge this gap by enabling communication in multiple languages. With the increasing importance of localization in the Indian market, there is a growing need for MT systems that can translate content from English to Indian languages and vice versa. Further, the availability of MT systems can make cross-border communication easier, faster, and more efficient, helping businesses to reach out to a wider audience and improve customer engagement. MT can also benef it government agencies, researchers, and individuals who need to communicate with people from different linguistic backgrounds. Therefore, the need for machine translation in India in the current global market scenario cannot be overstated, and efforts must be made to develop and improve MT systems to support Indian languages.

One of the significant institutions in India that have been working on Machine Translation research and development is the “Centre for Development of Advanced Computing” (CDAC) and its various centers, including the one in Pune, have been actively involved in developing MT methods for Indian languages. The CIS Department at the UoH and the IIIT in Hyderabad are also known for their research in MT for Indian languages. Additionally, the “Ministry of Communications and Information Technology” of the Government of India, via its TDIL Project, has supported the advancement of MT technologies for Indian languages. The Central Institute of Indian Languages in Mysore, the Amrita Vishwa

Vidyapeetham in Coimbatore and AUKBC in Chennai are other notable institutions that have contributed to MT research in India. The efforts of these institutions are crucial for addressing the challenges and opportunities of MT for Indian languages, and for promoting the use of local languages in various domains [4] [5].

The objective of our paper is to perform a survey on the existing methods of MT for the Indian languages including different challenges faced. In addition to that the key-findings from different surveys conducted on this topic are also highlighted along with current data-sources. In particular, the motivation is to pertain a set of entire research problems and findings regarding translating texts from one Indian language to another Indian language.

The contemporary and pertinent publications are searched from reputable databases such as IEEE Xplore, PubMed, and Google Scholar, using keywords such as ”machine translation,” ”Indian languages,” and ”recent developments.” Additionally, we explored proceedings of major conferences in natural language processing, including ACL and EMNLP, to capture the latest advancements.

This paper’s contribution is divided into nine subsequent sections. Section-II describes different MT approaches suitable for Indian languages. Section-III and Section-IV contain details discussions about MT-challenges and evaluation metrics for MT-Models respectively. Section-V highlights the timeline of important surveys conducted on Machine Translation in Indian languages for last 10 years. Section-VI helps to find datasets from different sources. On the unavailability of proper data-source some methods of constructing new data-sets are discussed in section-VII. Recent encouragement from the Indian government, as well as valuable contributions from renowned Institutions, are discussed in Section-VIII which draws the direction for future research. Section-IX summarizes our work in the conclusion.

2. Approaches to MT For Indian Languages

The field of MT comprises a range of techniques that are typically classified into different categories. Figure 2 displays several of these techniques and provides a timeline of their use over time.

A. Rule-based Machine Translation (RBMT)

RBMT relies on a set of human-created rules that specifies how a word or phrase in the source language should be translated into the target language. The rule set is determined by linguistic information such as morphology, vocabulary, syntax, phrase structure etc. RBMT works by matching the organization of the input sentence to that of the desired output sentence while preserving the original meaning of the input. After parsing the sentence in the source language, an transitional representation, like a parse tree or abstract representation, is generated. Figure 3 shows a general architecture of a RBMT system [6]. RBMT systems are again classified into Direct Translation, Transfer-Based Translation, and Interlingua categories based on the type of transitional representation they use.

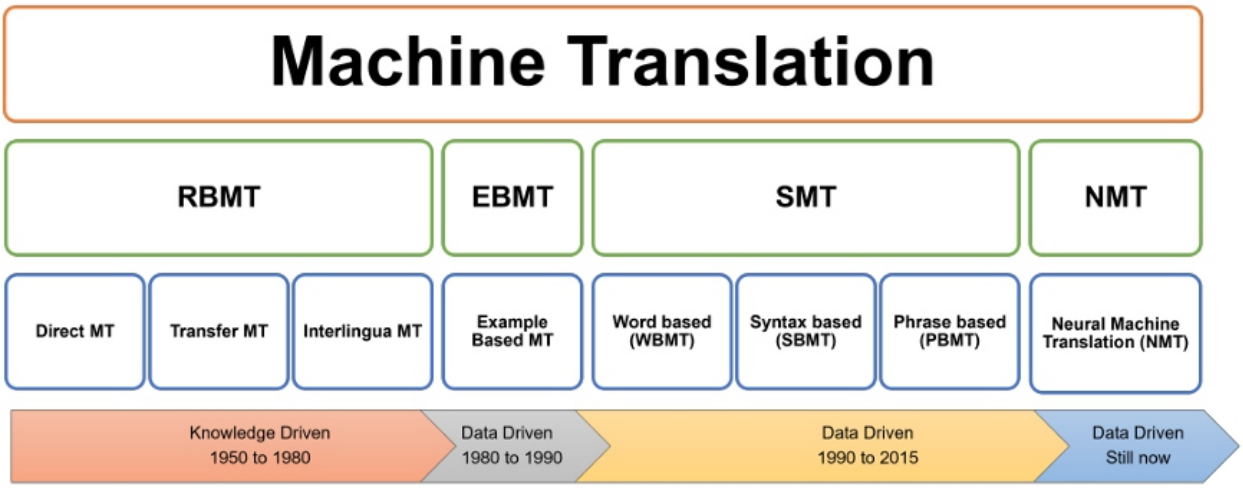


Figure 2. Approaches to Machine Translation with a Timeline

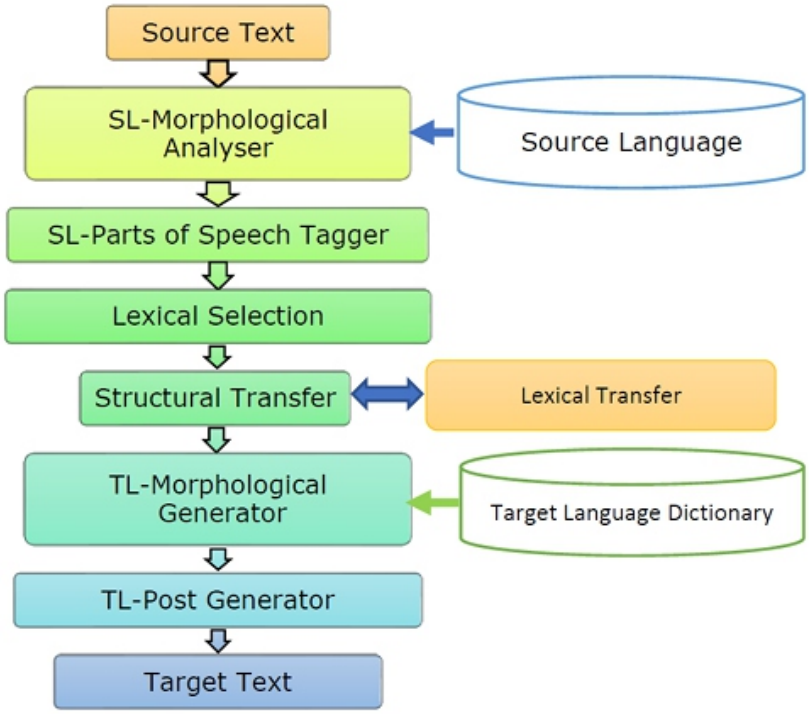


Figure 3. Architecture of RBMT approach

1) Direct Translation :

This simple method involves translating words directly from one language to another by using a bilingual dictionary, without considering the meaning or context of the source or target languages [7]. This approach can only handle one language pair at a time and is frequently unidirectional. From the late 1940s until the middle of the 1960s, the initial wave of machine translation was completely dependent on electronic or computer-readable dictionaries [8]. While this method works well for translating phrases, it is less successful when translating entire sentences.

2) Transfer Based Translation

Transfer-based machine translation is referred to as the second generation of MT's core (mid-1960s to 1980s). Transfer-based machine translation implies translating a sentence from the input language to a internal representation related with source language called as pivot language, and then from that pivot language to the target language. This approach allows for the use of more advanced translation techniques and takes into account the differences between the source and destination languages. However, it has the potential to introduce errors or lose meaning in the process of translating through a pivot language [8].

3) Interlingua Based

The Interlingua approach to MT prioritizes semantics and pragmatics above syntax. This method achieves the translation into two phases, the first of which involves converting the Source Language (SL) into an Interlingua (IL) form. The primary benefit of the Interlingua technique is that the SL analyzer and parser is not dependent on the Target Language (TL) generation and vice versa [9].

B. Example Based Machine Translation (EBMT)

Example-Based Machine Translation or EBMT is a translation methodology that uses a bilingual example database. By selecting pertinent instances from its example base, the EBMT system creates new translations. The target language translation is then created through processes of matching, alignment, and recombination [10].

C. Statistical Machine translation (SMT)

SMT method uses statistical models to learn patterns in a parallel corpus. A parallel corpus is a set of texts in two or more languages that are translations of each other. SMT system analyzes big amounts of bilingual parallel texts and forms the probabilistic model of how words, phrases, and sentences in one source language are related to the another target language. The statistical approach gained popularity recently due to the accessibility of bilingual parallel corpora and also the development of powerful statistical models and algorithms. The main benefit of SMT is that it can produce high-quality translations without the need for explicit linguistic knowledge or rules. Figure 4 shows the architecture of a typical SMT model. An SMT system aims to find the target sentence (comprising m words) y: y1, y2,...,ym, given a source sentence (comprising n words) x: x1, x2,...,xn,

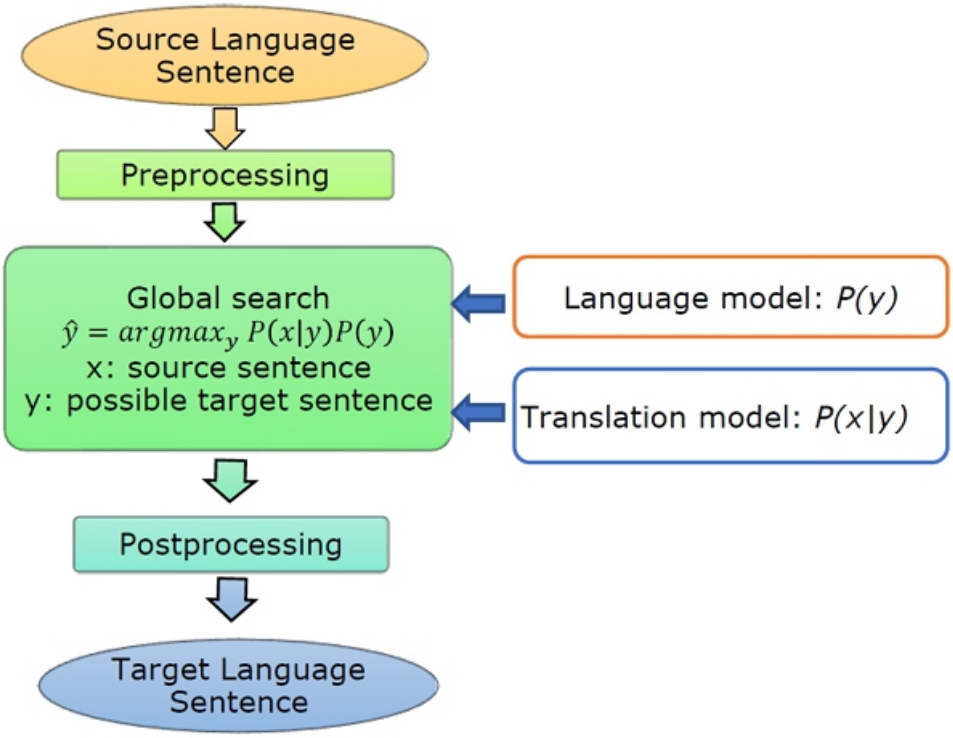


Figure 4. Architecture of a typical statistical machine translation system

such that the conditional probability p(y|x) is maximized. To achieve this, the Bayes rule is used.

$$\hat{y} = argmax_y P(y|x) = argmax_y P(x|y)P(y)..... \qquad (1)$$

P(y): a language model
$P(x|y)$: a translation model
$argmax_y$ = a decoder

The language model gets trained on the target language sentences (monolingual data) to maintain the fluency. Meanwhile, the translation model gets trained on parallel corpus of the source language and target language to identify lexical correspondences between them and their probabilities. A decoder is then used to combine the information from the language and translation models, and search for the best possible translation among all possible translations [11].

D. Neural Machine Translation (NMT)

NMT is the newest form of MT modeling that has succeeded in producing more accurate translations by exploiting huge amount of parallel text corpora. It relies on neural networks and deep learning techniques to create models based on existing reference translations. NMT requires a single sequence model, which leads to increased productivity. Using conditional probability modeling, NMT models the source phrase to the target sentence, producing a context vector c. Source phrase : x1, x2, x3,..., xm The target sentence : y1,y2,y3,...,yn

$$logP(y|x) = \sum_{m=1}^{n} logP(y_k|y_{k-1}, \ldots y_1, x, c) \qquad (2)$$

P(y|x) represents the likelihood of obtaining the target sentence words y given a source language word x, where c denotes the context of that specific word. The essence of NMT consists of two key elements: the ”encoder” and the ”decoder”. The input texts are transformed into a context vector (c) by the encoder , and subsequently, the decoder processes this vector to produce single word at a time for the output sentence with a length of m. Unlike other machine translation approaches, NMT requires minimal domain expertise [12]. The encoder-decoder model for NMT can be represented in a block diagram with figure 5.

1) Transformer

The attention-based NMT model which is also known as Transformer has revolutionized the field of machine translation for Indian languages. A transformer model was introduced by Google in 2017. It follows sequence-tosequence architecture involving encoders and decoders. Transformer models use an attention mechanism, which allows them to focus on the most relevant parts of the source sentence when generating the target sentence. This makes them more accurate and fluent than traditional machine translation models [13].

3. Challenges Of MT For Indian Languages:

Indian languages present a diversity of linguistic phenomena in terms of tense, gender, numbers, and other concepts. Due to structural and morphological complexity machine translation from English to Indian languages and vice versa is a challenging task. There are some challenges and problems faced during translation between Ils.

A. Syntactic Divergence

A fundamental structural distinction between English and Indian languages lies in the order of words in sentence. English sentences maintain the ’subject-verb-object’ order, whereas, the majority of Indian languages follow the ’subject-object-verb’ order. Certain Indian languages have a trait called free word order. Sense of prepositions in Indian languages are founded on specific symbolic conjunctive words however in English phrases, prepositions plays that role [14]. In English, prepositions come before the noun or pronoun they modify, whereas in the majority of Indian languages, they come after the noun or pronouns, which are also referred to as postpositions. Table-1 shows the divergence in word-order and use of prepositions in English and some Indian languages along with transliteration and word meaning [15].

B. Morphological Divergence

The field of morphology investigates the inner composition of words and their ability to take on unique shapes within different types of texts. The recognition, analysis, and description of morphemes as well as other linguistic constructions like words, affixes, and parts of speech are collectively referred to as ”morphology” in the study of language. The term ”morpheme” alludes to the lowest semantically significant item in a language. Words in the Indian language vary in terms of lemma, person, number, gender, case, tense, aspect, and modality. Languages with poor morphology typically use word order and syntax to convey various meanings. As a result, these languages have

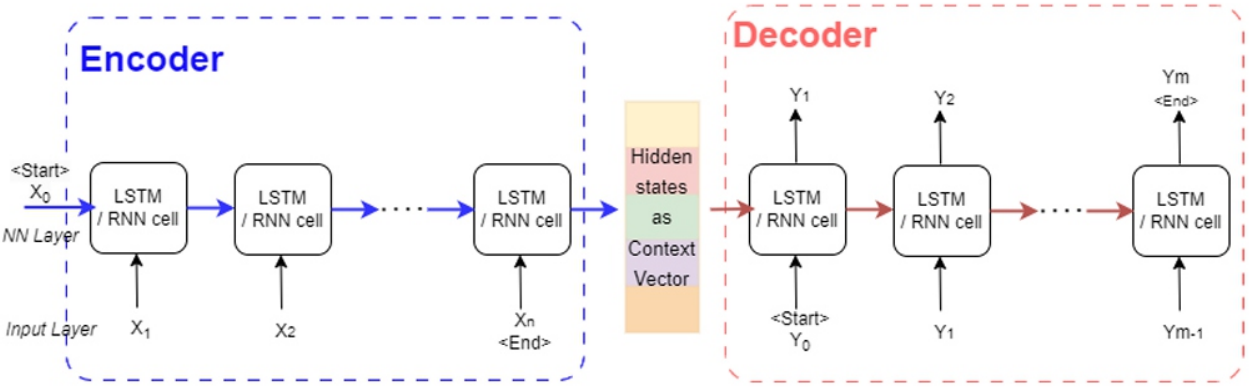


Figure 5. A General Encoder-Decoder Model

English sentence	Hindi	Bangla	Marathi
I have a pen	मेरे पास एक कलम है mere (I) paas (pp) ek (a) kalam(pen) hai (have)	আমার একটি কলম আছে Āmāra (I) ēkaṭi (a) kalama (pen) āchē (have)	माझ्याकडे पेन आहे Mājhyākaḍē (I) pēna (pen) āhē (have)
The cat is on the table	बिल्ली मेज पर है billee (the cat) mej (table) par (on) hai (is)	বিড়াল টেবিলের উপর Biṛāla(the cat) ṭēbilēra (table) upara (on) **verb is silent	मांजर टेबलावर आहे Māñjara(cat) ṭēbalāvara(on table) āhē (is)

Figure 6. Word-order divergence among Indian languages

a smaller lexicon than languages with a rich morphological structure. Richer languages have more nuanced words that accurately communicate various meanings, which increases the language's complexity. Hebrew, Turkish, Dravidian languages, and other languages are thought to be morphologically rich, whereas English, Mandarin, and other languages are thought to be morphologically poor. Due to a bigger vocabulary, sparser data, and increased complexity, morphologically rich languages are more difficult for neural networks to model than poor ones. The Stochastic Morph Analyzer (SMA) is a Morph Analyzer that forecasts the morph information using machine learning [16] [17]. In India, Dravidian languages such as Telugu and Tamil exhibit greater morphological complexity compared to Indo-Aryan languages like Hindi, Punjabi, and Gujarati. Translating text into Dravidian languages like Telugu, Tamil, and Malayalam often yields lower BLEU scores, whereas translations into Indo-Aryan languages like Hindi, Gujarati, Punjabi, and Bengali tend to achieve relatively higher BLEU scores. Alarger number of distinct words can be found in the richer languages within a multilingual parallel corpus. Morphological complexity can be measured by Type-Token ratio. Here is the increasing order of morphological complexity for different languages:- Hindi<Punjabi<Gujarati<Tamil<Telugu [18].

C. Data scarcity

Building of Corpus can be expensive for users with limited resources. When the word order is significantly diverse between two languages, statistical machine translation struggles. NMT does not

come up to the mark for morphologically diverse languages.

D. Interpreting the intentions of speakers is challenging

Depending on the speaker's aim (such as sarcasm, sentiment, metaphor, etc.), phrases or words might have many interpretations.

E. Code-mixed language

Processing code-mixed language is difficult because users often utilize numerous languages in a single statement or phrase. E.g.: User tweet : "Hi friends, keyse ho? Ayo chill kare."

F. Idioms

Sometimes idioms may not be interpreted idiomatically. Indian regional languages are rich with idioms.

4. Evaluation Metrics of MT-Algorithms

To measure the goodness of a MT-model several metrics such as BLEU, METEOR, ROUGE, TER, NIST etc. are available for automatic evaluation. Evaluation metrics can be categorized into 2 types, Intrinsic Evaluation and Extrinsic Evaluation.Both intrinsic and extrinsic evaluation metrics are focused on the performance of the final objective, which is the performance of the NLP component on the entire application, whereas intrinsic evaluation metrics are more concerned with intermediate objectives, such as how well an NLP component performs on a specified subtask.
We discussed some common intrinsic evaluation metrics used for MT systems.

A. Bilingual Evaluation Understudy (BLEU)

The BLEU metric calculates the score by comparing n - grams of the candidate translation of text to one or more ngrams reference translations. The BLEU metric ranges from 0 to 1. A score of 1.0 denotes a perfect match, whereas a score of 0.0 denotes a perfect mismatch. Sometimes BLEU score is expressed as a percentage rather than a decimal between 0 and 1. The following interpretation of BLEU scores (expressed as percentages rather than decimals) is followed in general [19].
The provided color gradient can serve as a broad representation of the BLEU score on a scale. It is the most widely accepted, inexpensive and easily understandable metric.

B. Metric for Evaluation of Translation with Explicit OR-dering (METEOR)

METEOR is based on the unigram matching and calculated by the harmonic mean of precision and recall. The recall is higher weighted than precision. It overcomes some of the drawbacks of the BLEU score, as because it can perform stemming- and synonymity matching, as well as standard exact word-matching [20]. This is a perfect metric for Machine translation. Once the final alignment is computed, the score of Unigram precision P and Unigram Recall R is calculated as:

$$P = \frac{m}{w_t}, \qquad R = \frac{m}{w_r} \qquad (3)$$

where m = no. of unigrams in the observed translation that are also available in the reference translation, wt = no of unigrams in the observed translations, wr = no of unigrams in the reference translations. The harmonic mean (F) is calculated as :

$$F_{mean} = \frac{10PR}{(R + 9P)} \qquad (4)$$

C. Recall-Oriented Understudy for Gisting Evaluation (ROUGE)

ROUGE basically measures the "recall" or overlap, between the generated text and the reference summaries, providing a quantitative measure of the content overlap and effectiveness of the generated output. It is used in machine translation projects to assess the quality of the text that is produced [21].

D. Translation Error Rate (TER)

TER quantifies the number of editing operations needed to align a translated segment with a reference translation. TER score ranges from 0% to 100%. The quality of the translation improves with decreasing TER scores. A higher BLEU or METEOR score, on the other hand, indicates better translation quality. A better MT system achieves higher BLEU scores with lower CDER, TER and PER scores [21] [22].

E. National Institute for Standards and Technology (NIST) from US

It is based on BLEU metric with some features. The n-gram precision calculation is differently taken. In contrast to BLEU, which assigns equal weight to all n-grams, NIST takes into account the relevance of each n-gram. It assigns higher weight to n-grams that are considered less likely to occur [23]. Metrics for automatic evaluation are quick, tuneable, affordable, and require less human labour. But these automatic evaluation metrics are not adequate for evaluating MT systems in Indian languages. Due to the many intricacies involved with Indian languages, they will not generate reliable results, but same measures produce excellent evaluation results for Non-Indic western languages.

For evaluating the quality of translated phrases, human evaluation metrics are preferred for particularly morphologically rich languages, despite being time-consuming and costly. Human evaluation entails bilingual expertise in both the source and target languages, offering a level of consistency often deemed superior to automatic translation assessments [21].

5. Recent MTResearch For Indian Languages

In this section we highlight important research work done for Indian languages with a focus on low-resource languages. Jindal et al. 2018 used SMT based MT model for translation between English and Punjabi using three sets of parallel-sentence corpus achieving 0.8767 BLUE score [24].

Mahata et al. 2018 implemented RNN encoder-decoder architecture to improve the quality of translation done by traditional SMT. English-Hindi parallel corpus from MTIL2017 was used as dataset to analyse the scores of phrase-pairs by a comparative experiment between two models. It was found that SMT performs fine for long sentences and NMT performs well for short sentences [25]. Pathak et al. 2019 exploited OpenNMT system architecture for English to Punjabi, Tamil, and Hindi languages. They observed the betterment of performance of NMT model with the growth in the training data and length of test sentences [26].

Shah et al. 2019 constructed an Attention-based Encoder-Decoder model featuring 128 LSTM cells and 2 layers. Their experimentation involved a self-created

TABLE I. Interpretation of BLEU scores in percentage

BLEU Score	Interpreted as
Less than 10	Not useful
10 to 19	It's hard to obtain the meaning
20 to 29	The sense is clear, but it has large grammatical errors
30 to 40	Translations quality is good
40 to 50	Translations quality is high
50 to 60	Very high-quality, acceptable, and smooth translations
Greater than 60	Quality is quite acceptable than human-efforts

Figure 7. BLEU Score Table

bilingual dataset encompassing English and Gujarati for translation purposes. Notably, the model demonstrated a commendable BLEU score of 40.33 during the testing phase [27].

Bansal et al. 2020 proposed a method for enhancing NMT and handling Out-of-Vocabulary (OOV) words by combining word level and character level attention information. The method used two attention mechanisms, with the first mechanism employing Gated Recurrent Unit (GRU) character-level attention and the second mechanism utilizing word-level attention. The encoder simultaneously encodes information from both character and word levels, while the decoder decodes based on word-level attention only. The authors achieved BLEU score as 27.65 and WER 30.17 for English-Hindi language pair [28].

Tatwadarshi et al. 2020 exposed the necessity of MT systems in the Indian perspective because more than 50% of the data generated online is in English which only 12% of Indian people know. The Neural Machine Translation system developed by Google and Facebook are less effective for syntactically complex languages like Indian languages. They have primarily prioritized parallel translation over contextual accuracy of the sentence. The author proposed a conceptual framework by combining document and sentence-level contextual information and an Indian Language-English contextual dictionary fed together with a bi-lingual parallel corpus to the NMT model. The proposed system was expected to address the specific challenges of Indian MT system [29].

Dewangan et al. 2021 worked for Indian Language NMT using one of the popular subword methods i.e., BPE based NMTmodel. They used ILCI dataset to derive BLEU scores for different pairs of languages . The authors proposed a data augmentation technique which combined NMT and SMT [30]. Laskar et al. 2021 participated in 'Workshop on Asian Translation 2021' multimodal translation task of English to Hindi. An investigation was done for phrase pairs through data augmentation technique in multimodal and text-only NMTsystems. The results were evaluated by BLUE, Rankbased Intuitive Bilingual Evaluation (RIBES), and Adequacy Fluency Metrics (AMFM) which scored better than the previous works [31].

A Chowdhury et al. 2022 used Transfer Learning approach for translation between a low-resource Indian language called Lambani and other Indian languages. The BLEU score was improved when the TL was

used and the authors have observed that freezing the initial layers of the TL model improved the BLUE score further [32].

As part of the AI4Bharat Initiative, Divyanshu et al.2020 developed ”IndicBERT,” a multilingual pre-trained model based on the compact ALBERT architecture [33]. The word-embedding methods employed are suitable for morphologically rich languages. The model underwent pretraining on a monolingual corpus containing 12 Indian languages and 9 billion tokens. Additionally, the authors have made significant contributions by providing several NLP datasets and models for research on Indian languages as open source [34].

Jay Gala et al.2023 have developed a translation model for 22 Indian languages named IndicTrans2. Under this project, the authors have released different variants of indicindic model intending to improve the quality of direct translation. Their MT models use English as pivot language, hence there are scopes of further improvement [35]. Some important points from past recent surveys on Machine translation in Indian languages have been summarized in Table II.

6. Availability of Dataset

This section discusses some open-source datasets for the automatic translation between Indian languages. A parallel

TABLE II. Key points of past few surveys on ML in Indian Languages

Year	Key observations and limitations	Ref No.
2015	Transferred-Based approach is more flexible. Most of the MT research work has been done in Aryan languages. Dravidian languages are yet to be explored.	[36]
2018	Automatic Performance metrics for MT algorithms are not adequate. Human Evaluation metrics are suitable for Indian Languages. Existing systems performance is not satisfactory.	[37]
2018	Machine Translations are carried out between English and Indian languages, with the exception of Google Translator.	[38]
2019	The USA leads the world in MT research followed by Japan, China. India is still now in the infancy stage of MT due to it’s language-diversity. MT-research can be improved by govt. policies for the benefit of society.	[39]
2019	Low-resource languages should be focused more for future studies in terms of the availability of data-sources, translation methods, and challenges for translation.	[40]
2020	Despite having a vast body of ancient Indian literature and science, the Sanskrit language has received very little attention. Hybrid and NMT methods show better performance as compared to other techniques.	[41]
2021	SMT performs well for translation among Indo-Aryan family, but is poor for Dravidian family.	[30]

text corpus is comprised of pairs of sentences, one in source language and another in target language and the meaning of the both sentences are same.

A. The EMILLE Corpus

The EMILLE (Enabling Minority Language Engineering) Corpus was created by collaboration among the CIIL, Mysore, India, Lancaster University, UK. The corpus is made up of three parts: parallel, monolingual, and annotated corpora. The fourteen monolingual corpora for fourteen south Asian languages are Bengali, Assamese, Hindi, Gujarati, Malayalam, Telegu, Kannada, Tamil, Kashmiri, Punjabi, Marathi, Oriya, Sinhala, and Urdu. They contain written and spoken data which is provided without charge for use in exclusively non-commercial research [42].

B. IJCNLP-2008 data set

This dataset was developed for the Named Entity Recognition (NER) challenge in a workshop hosted by IIIT, Hyderabad about NER for South East Asian languages. It included Hindi, Bengali, Oriya, Telugu, and Urdu databases [43].

C. Tatoeba

The Tab-delimited Bilingual Sentence Pairs datasets are created by Tatoeba project by compiling statements from many languages. They paid particular attention to the creation of numerous linguistic datasets that included translations of sentences in various low-resource languages. Many low-resource language to English translation can be done using this dataset. The tab key serves as a line between the original and translated sentences. Each dataset contains at least 100 sentences and their translations [44]. Table III highlights a few sample snapshots of the accessible data sources.

D. Anuvaad

It is an open-source platform for translating court papers at scale in the judicial sector. Supreme Courts of India (SUVAS) and Bangladesh (Supreme Court) have separate Anuvaad instances deployed (Amar Vasha). Now Anuvaad have high quality NMT models for nine Indian languages [45] [46].

E. AI4Bharat

AI4Bharat is the recent initiative of IIT Madras. It aims on building a rich open-source language AI system for Indian languages, including datasets, models, and applications. Samanantar is an extensive parallel corpus collection for Indic languages that is accessible to the public [47] [48].

F. Mann ki Baat

"Mann Ki Baat"– is a monthly program of All India Radio in which the Prime Minister of India speaks and addresses the citizens in Hindi language. Later the speech is converted to different other Indian languages. The Textual Data or Parallel corpus for Indian languages can be mined
from multilingual articles called "CVIT Mann Ki Baat" [49] [50] [51].

G. Universal Language Contribution API (ULCA)

ULCAisastandard API and open scalable data platform under Bhashini which supports various types of datasets and models for Indian languages. Bhashini serves as an artificial intelligence tool strategically created to overcome language barriers prevalent among the various languages spoken

TABLE III. Example dataset snap of sentence pairs from the Tatoeba Project

This data is from tatoeba project
Link : "http://tatoeba.org/files/downloads/sentences_detailed.csv"
Date of this file: 2023-09-06

Bengali to English	
Nobody was home.	কেউ বাড়ি ছিলো না।
Nothing changed.	কিছুই পাল্টালো না।
Nothing's there.	ওখানে কিছুই নেই।
Please hurry up.	একটু তারাতারি করুন।
Please hurry up.	একটু তারাতারি করো।
Kannada to English	
I don't know what came over me.	ನನಗೆ ಏನಾಯಿತು ಅಂತ ನನಗೇನೆ ಗೊತ್ತಿಲ್ಲ.
Do you think it means something?	ಅದಕ್ಕೆ ಅರ್ಥ ಇದೆ ಎಂದು ನಿಮಗೆ ಅನಿಸುತ್ತಾ?
I'm glad you guys could make it.	ನೀವೆಲ್ಲ ಬಂದಿದ್ದು ನನಗೆ ತುಂಬಾ ಸಂತೋಷ.
I'm sorry I missed your concert.	ಕ್ಷಮಿಸಿ ನಿಮ್ಮ ಕಛೇರಿಗೆ ಬರುವದಕ್ಕೆ ಆಗಲಿಲ್ಲ.
Some animals are afraid of fire.	ಕೆಲವೊಂದು ಪ್ರಾಣಿಗಳು ಬೆಂಕಿಗೆ ಹೆದರುತ್ತವೆ.
Tell us what you know about Tom.	ಟಾಮ್ ಬಗ್ಗೆ ಏನೆಲ್ಲಾ ಗೊತ್ತೋ ನಮಗೆ ಹೇಳಿ.

across India. This tool provides instantaneous translation capabilities and empowers developers to utilize an opensource language database for constructing tools, applications, and services in regional languages. Through the online crowd-sourcing platform 'Bhashadaan' the contributors can take part into four programs– 'Suno India', 'Likho India', 'Bolo India' and 'Dekho India'. The prime minister of India inaugurated Bhashini in 2022 at Gujarat [52].

7. Initiative of Constructing Parallel Corpora

Indic languages often have an abundance of monolingual corpora but a scarcity of parallel corpora, making it challenging to apply machine-engineered techniques for dataset creation. The following are some of the reasons that make the creating parallel data a difficult task:

1) Many data are not in digital format. Some of them are either in PDF files or in image format.
2) Texts are not in Unicode. they use proprietary font formats.
3) Many datasets are not in format that can be directly used for MT. The incomplete sentence, invalid character sequence, spell errors, mixed with other language etc. create immature dataset for machine translation.

Thus, in order to construct machine translation systems for Indic languages, it is imperative to either create synthetic parallel corpora or use language models in the system's training.

Steps to create Bilingual Parallel corpora:

1) Selection of the Source and the Target Language
1) Selection of the Source and the Target Language
2) Collection of source and target texts from books, newspapers, websites and other documents.
3) Preprocessing: cleaning errors, formatting, and extraneous characters.
4) Alignment of source and their corresponding target texts by different automated tools (Bluealign, Giza++, Ugarit) [53]

5) Annotation: After alignment, the parallel corpus needs to be annotated with metadata such as a sentence or phrase-level information, part-of-speech tags, named entities, and other linguistic features.
6) Quality control: Finally, the parallel corpus needs to be checked for quality control to ensure accuracy and consistency in translations.

Under the project MTIL-2017 Shared Task an initiative was taken by M. Anand Kumar et. al to develop parallel corpora between English and Indian languages in September 2017 by conducting a shared task among 29 teams of people. The team worked with Hindi, Tamil, Malayalam, and Punjabi languages and employed Neural Network based system. The output evaluation was done by human beings [54]. Philip et al. [55] built a standard NMT system, a retrieval module, and an alignment module make up the iterative alignment pipeline. This pipeline is used to interact with publicly accessible websites, such as government news releases. As more articles are published to PIB and additional tools are put in place to gather more sentences, the corpus will undoubtedly grow in size.

8. Indian Govt. Encouragement and Future Scope Of MT

The following 22 languages are listed in the Constitution's Eighth Schedule. Initially 14 languages were listed as : 1) Assamese, 2) Bengali, 3) Gujarati, 4) Hindi, 5) Kannada, 6) Kashmiri, 7) Malayalam, 8) Marathi, 9) Oriya, 10) Punjabi, 11) Tamil 12) Telugu, 13) Urdu and 14) Sanskrit. Later on more 8 languages like Bodo, Dogri, Konkani, Maithili, Manipuri, Nepali, Santali and Sindhi were included in the list [2].
To lower the barriers to communication, various organisations in India are supporting the adoption and integration of MT technologies and programmes. India is positioned to experience tremendous growth in the international IT sector with the launch of the government's "Digital India" plan. Initiatives like Digital India promise to provide plenty of chances for national and international businesses to broaden and deepen their penetration into Indian markets.

A. CIIL

In Mysore, Karnataka, the Central Institute of Indian Languages (CIIL) was established to oversee the development of Indian languages [56]. The CIIL, the Ministry of Human Resource Development's (MHRD) nodal organisation is responsible for the promotion and preservation of Indian languages. Some newer projects of the CIIL are:
- New Language Survey of India (NLSI).
- LDC-IL.
- National Translation Service.
- Development and promotion of minor Indian languages.
- Development of Pali.
- National Testing Mission.

B. ILCI

The Indian Languages Corpora Initiative (ILCI), a massive effort started by the Indian government, aims to compile parallel annotated corpora in each of the 17 languages listed in the Indian Constitution. ILCI project aims to provide a common language platform by developing parallel annotated corpora in the tourism and health sectors in 11 Indian languages, with Hindi serving as the source language. The project's primary goal is to create an annotated parallel corpus from source Hindi to Indian languages with English [30].

C. C-DAC

C-DAC is a research and development organization that operates under the MeitY of the Government of India. Its mission is to develop tools for multilingual translation and methods to bridge the gap between Indian languages due to the country's multilingual nature. C-DAC provides users with access to these resources for their research projects. Additionally, it offers dictionaries and corpora for Indian languages, among other resources [57].

D. TDIL

The Government of India's Meity initiated the Technology Development for Indian Languages (TDIL) Program. The primary objective is to facilitate the creation and accessibility of multilingual knowledge resources. The program also strives to develop tools and techniques for information processing, fostering human-machine interaction devoid of language barriers. An additional goal involves the integration of these advancements to craft innovative user products and services. The program also actively participates in national and international standardization organizations such as UNICODE, ISO, the W3C, and BIS to promote language technology standardization and ensure appropriate description of Indian languages in current and future standards [4].

Though research in MT for Indian languages has grown tremendously during the past decade, certain areas are yet to be explored such as Code-mixed IL processing, Opinion mining, sarcasm translation, idioms extraction for Indian languages.

9. Conclusion

In this paper, we projected some light on the previous works related to Machine translation for Indian languages by keeping in mind the rising demand for research in the multilingual translation process of India. We presented a systematic as well as comprehensive review of the different methods of MT for Indic languages and the challenges faced by other researchers in this regard. To establish a rigorous evaluation process, this review engages in an indepth exploration of various evaluation metrics employed in the domain of machine translation. We have also included the most recent references of a detailed source of available datasets, The importance of parallel corpora is crucial for MT research in India. Yet, it has been noted that there are still no suitable techniques for producing parallel corpora datasets. We also provided some insight into earlier attempts made in this area. Finally, there are many opportunities for machine translation research in India. Thanks to Indian government's strong encouragement and assistance through the Digital India program.

References

[1] R. A. Sinhal, Dept. of CSE Shri Ramdeobaba College of Engineering and Management Nagpur, and K. O. Gupta, "Machine translation approaches and design aspects," IOSR J. Comput. Eng., vol. 16, no. 1, pp. 22–25, 2014.

[2] Department of Higher Education, "Language Education," https://www.education.gov.in/sites/upload files/mhrd/files/upload document/languagebr.pdf, Data Collected on 29-11-2023.

[3] Ministry of Home Affairs, Government of India, "Report of the Committee of Parliament on Official Language," https://rajbhasha. gov.in/sites/default/files/cpol7threporteng.pdf, 2005.

[4] Ministry of Electronics Information Technology, Govt. of India, "Final draft standard on machine translation acceptance. Version 4.0." https://tdil-dc.in/index.php?option=com rff article& task=view-article&article id=11&lang=en, Data Collected on 2911-23

[5] G. N. Jha, "The tdil program and the indian language corpora initiative (ilci)," in Proceedings of the International Conference on Language Resources and Evaluation (LREC), 2010.

[6] N. Bhadwal, P. Agrawal, and V. Madaan, "A machine translation system from hindi to sanskrit language using rule based approach," Scalable Computing: Practice and Experience, vol. 21, pp. 543–554, 2020.

[7] R. Sankaravelayuthan and G. Vasuki, "English to tamil machine translation system using corpus," http://www.languageinindia.com/, 2013, unpublished Manuscript. parallel

[8] V. S.-C. Yang, "Electronic dictionaries in machine translation," in Encyclopedia of Library and Information Science, A. Kent and et al., Eds. New York: Marcel Dekker, 1991, vol. 48, no. Suppl. 11, pp. 74–92.

[9] A. Chatterjee, Elements of Information Organization and Dissemination, 2017, doi:10.1016/B978-0-08-102025-8.00025-9.

[10] K. M. Anwarus Salam, M. Khan, and T. Nishino, "Example based english-bengali machine translation using wordnet," in Conference Proceedings- Centre for Research on Bangla Language Processing. BRAC University, 2010.

[11] M. Mumin, M. Hanif, M. Iqbal, and M. J. Islam, "shu-torjoma: An englishbangla statistical machine translation system," Journal of Computer Science, vol. 15, pp. 1022–1039, 08 2019.

[12] S. Sharma and M. Diwakar, "Machine translation for indian languages utilizing recurrent neural networks and attention," in Distributed Computing and Optimization Techniques, ser. Lecture Notes in Electrical Engineering, S. Majhi, R. P. d. Prado, and C. Dasanapura Nanjundaiah, Eds., vol. 903. Springer, Singapore, 2022.

[13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2023.

[14] M. D. Okpor, "Machine translation approaches: Issues and challenges," International Journal of Computer Science Issues (IJCSI), vol. 11, no. 5, p. 159, 2014.

[15] R. N. Patel, P. B. Pimpale, and M. Sasikumar, "Machine translation in indian languages: Challenges and resolution," Journal of Intelligent Systems, vol. 28, no. 3, pp. 437–445, 2019.

[16] S. Srirampur, R. Chandibhamar, and R. Mamidi, "Statistical morph analyzer (SMA++) for Indian languages," in Proceedings of the First Workshop on Applying NLP Tools to Similar Languages, Varieties and Dialects. Dublin, Ireland: Association for Computational Linguistics and Dublin City University, Aug. 2014, pp. 103–109.

[17] A. Bharati, R. Sangal, S. Bendre, P. Kumar, and Aishwarya, "Unsupervised improvement of morphological analyzer for inflectionally rich languages." in Proceedings of the Sixth Natural Language Processing Pacific Rim Symposium, 01 2001, pp. 685–692.

[18] A. Tanwar and P. Majumder, "Translating morphologically rich indian languages under zero-resource conditions," ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 19, no. 6, p. Article 85, November 2020.

[19] G. Cloud. (2024) Automl api documentation. Accessed: 24 Feb 2024. [Online]. Available: https://cloud.google.com/translate/ automl/docs/evaluate

[20] S. Banerjee and A. Lavie, "METEOR: an automatic metric for MT evaluation with improved correlation with human judgments," in Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization, 2005, pp. 65–72.

[21] A. Kandimalla, P. Lohar, K. Maji, and A. Way, "Improving english-to-indian language neural machine translation systems," Information, vol. 13, p. 245, 2022. [Online]. Available: https: //www.mdpi.com/2078-2489/13/5/245

[22] P. Madaan and F. Sadat, "Multilingual neural machine translation involving indian languages,"

Data: Resources and Evaluation. Marseille, France: European Language Resources Association (ELRA), 2020, pp. 29–32.

[23] G. Doddington, "Automatic evaluation of machine translation quality using n-gram co-occurrence statistics," in Proceedings of the Second International Conference on Human Language Technology Research. San Diego, California: Morgan Kaufmann Publishers Inc., 2002, pp. 138–145.

[24] S. Jindal, V. Goyal, and J. S. Bhullar, "English to punjabi statistical machine translation using moses (corpus based)," Journal of Statistics and Management Systems, vol. 21, no. 4, pp. 553–560, 2018.

[25] S. K.Mahata, D. Das, and S. Bandyopadhyay, "MTIL2017: Machine translation using recurrent neural network on statistical machine translation," Journal of Intelligent Systems, vol. 28, no. 3, pp. 447453, 2019.

[26] A. Pathak and P. Pakray, "Neural machine translation for indian languages," Journal of Intelligent Systems, vol. 28, no. 3, pp. 465477, 2019.

[27] P. Shah and V. Bakrola, "Neural machine translation system of indic languages- an attention based approach," in 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), 2019, pp. 1–5.

[28] M. Bansal and D. Lobiyal, "Word-character hybrid machine translation model," in 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2020, pp. 270–274.

[29] T. P. Nagarhalli, V. Vaze, and N. K. Rana, "A novel framework for neural machine translation of indian-english languages," in 2020

International Conference on Inventive Computation Technologies (ICICT), 2020, pp. 676–682.

[30] S. Dewangan, S. Alva, and e. a. Joshi, Nisarg, "Experience of neural machine translation between indian languages," Machine Translation, vol. 35, pp. 71–99, 2021.

[31] S. R. Laskar, A. F. U. R. Khilji, D. Kaushik, P. Pakray, and S. Bandyopadhyay, "Improved English to Hindi multimodal neural machine translation," in Proceedings of the 8th Workshop on Asian Translation (WAT2021). Online: Association for Computational Linguistics, Aug. 2021, pp. 155–160.

[32] A. Chowdhury, D. K. T., S. V. K, and S. R. Mahadeva Prasanna, "Machine translation for a very low-resource language- layer freezing approach on transfer learning," in Proceedings of the Fifth Workshop on Technologies for Machine Translation of LowResource Languages (LoResMT 2022). Gyeongju, Republic of Korea: Association for Computational Linguistics, 2022, pp. 48–55.

[33] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "Albert: A lite bert for self-supervised learning of language representations," 2020.

[34] D. Kakwani, A. Kunchukuttan, S. Golla, G. N.C., A. Bhattacharyya, M. M. Khapra, and P. Kumar, "IndicNLPSuite: Monolingual Corpora, Evaluation Benchmarks and Pre-trained Multilingual Language Models for Indian Languages," in Findings of EMNLP, 2020.

[35] J. Gala, P. A. Chitale, R. AK, V. Gumma, S. Doddapaneni, A. Kumar, J. Nawale, A. Sujatha, R. Puduppully, V. Raghavan, P. Kumar, M. M. Khapra, R. Dabre, and A. Kunchukuttan, "Indictrans2: Towards high-quality and accessible machine translation models for all 22 scheduled indian languages," 2023.

[36] S. Saini and V. Sahula, "A survey of machine translation techniques and systems for indian languages," in 2015 IEEE International Conference on Computational Intelligence & Communication Technology, 2015

[37] D. Chopra, N. Joshi, and I. Mathur, "A review on machine translation in indian languages," Eng. Technol. Appl. Sci. Res., vol. 8, no. 5, pp. 3475–3478, October 2018.

[38] N. Kharate, "Survey of machine translation for indian languages to english and its approaches," 3,

3, pp. 613–623, 2018.

[39] B. M. Gupta and S. Dhawan, "Machine translation research: A scientometric assessment of global publications output during 200716," DESIDOC Journal of Library & Information Technology, vol. 39, pp. 31–38, 2019.

[40] B. S. Harish and R. K. Rangan, "A comprehensive survey on indian regional language processing," SN Applied Sciences, vol. 2, 2020.

[41] M. Singh, R. Kumar, and I. Chana, "Machine translation systems for indian languages: Review of modelling techniques, challenges, open issues and future research directions," Archives of Computational Methods in Engineering, 2020.

[42] "Emille dataset," https://www.lancaster.ac.uk/fass/projects/corpus/ emille/, 2003.

[43] "Ijcnlp dataset," http://ltrc.iiit.ac.in/ner-ssea-08/, 2008.

[44] "Tab-delimited bilingual sentence pairs," https://www.manythings. org/anki/, 2015.

[45] "Anuvad dataset," https://www.anuvaad.org/.

[46] "Anuvad parallel corpus," https://github.com/project-anuvaad/ anuvaad-parallel-corpus.

[47] "Samanantar parallel corpus," https://ai4bharat.org/samanantar.

[48] G. Ramesh, S. Doddapaneni, A. Bheemaraj, M. Jobanputra, A. Raghavan, A. Sharma, S. Sahoo, H. Diddee, M. J, D. Kakwani, N. Kumar, A. Pradeep, S. Nagaraj, K. Deepak, V. Raghavan, A. Kunchukuttan, P. Kumar, and M. S. Khapra, "Samanantar: The largest publicly available parallel corpora collection for 11 indic languages," Transactions of the Association for Computational Linguistics, vol. 10, pp. 145–162, 2022.

[49] "Mann ki baat," Accesed on : 30/12/2023. [Online]. Available: https://www.pmindia.gov.in/en/mann-ki-baat/

[50] A. Project, "Indicnlp catalog," https://github.com/AI4Bharat/ indicnlp catalog.

[51] T. Tiwari. (Year of access) Pm india mann ki baat. [Online]. Available: https://www.kaggle.com/datasets/taruntiwarihp/ pm-india-mann-ki-baat

[52] "Ulca-bhashini," https://bhashini.gov.in/ulca.

[53] A. Sati, "Word alignment using giza++ and cygwin on windows," International Journal of Engineering Research & Technology (IJERT), vol. 02, no. 05, May 2013.

[54] M. A. Kumar, B. Premjith, S. Singh, S. Rajendran, and K. P. Soman, "An overview of the shared task on machine translation in indian languages (mtil)– 2017," Journal of Intelligent Systems, vol. 28, no. 3, pp. 455–464, 2019.

[55] J. Philip, S. Siripragada, V. P. Namboodiri, and C. V. Jawahar, "Revisiting low resource status of indian languages in machine translation," in Proceedings of the 3rd ACM India Joint International Conference on Data Science & Management of Data (8th ACM IKDD CODS & 26th COMAD), January 2021.

[56] "Central institute of indian languages (ciil)," https://www.ciil.org/ default.aspx.

[57] "Cdac dataset," https://www.cdac.in/index.aspx?id=products services, 2003.

Sudeshna Sani is a research scholar at KLEF (K L Deemed to be University), Vaddeswaram, Andhra Pradesh, India. She received her Diploma in Computer Science and Technology from WBSCTE, AMIE (CSE) from the IEI (India), Kolkata and thereafter M.Tech(CSE) from MAKAUT, West Bengal, India in the year of 2005, 2017 and 2019 respectively. Presently, she is working as an Assistant Professor at Woxsen University, Hyderabad. She has 5 years of experience of teaching various courses of Computer Science. She also carries more 8 years' of experience as Lab-Technician in engineering colleges exploring various technical aspects of Computer Science. She is a lifetime member of the ISTE, and IEI. Her research interest is focused on NLP, Machine Translation, speech and text processing through machine learning and deep learning-based systems.

Dr. Samudravijaya K is a Professor at KLEF (K L Deemed to be University), Vaddeswaram, Andhra Pradesh, India, since January 2022. He carried out research in the area of Spoken Language Processing at Tata Institute of Fundamental Research, Mumbai for 3 decades. Later, he was a Visiting Faculty at IIT Guwahati during 2016 2020 and an Adjunct Faculty at IIT Dharwad during 2021-2022. He received awards such as 'Best Ph.D. Thesis Award' during the year 1986, UNDP Fellowship for research at Carnegie Mellon University, USA in 1988, Sir C V Raman Award by the Acoustic Society of India in 2003 etc. His research interests include Speech and speaker recognition. Voice enabled information access systems and Machine Learning. He has been teaching various courses in these areas. He is guiding research scholars in the areas of Speech Processing in addition to contributing to several sponsored projects.

Dr. Suryakanth V Gangashetty is a Professor in the Department of Computer Science and Engineering at KLEF (KL University) Vaddeswaram, Guntur District, Andhra Pradesh, India. He completed his PhD (in Neural Network Models for Recognition of Consonant-Vowel Units of Speech in Multiple Languages) from IIT Madras in 2005. Before joining KLEF Vaddeswaram , he

worked as a member of faculty at IIIT Hy
derabad, Telangana, from 2006 to 2020. Previously he has worked as a Senior Project Officer at Speech and Vision Laboratory, IIT Madras. He has worked as a member of faculty at BIET Davangere Karnataka, from 1991 to 1999. He has also worked as a visiting research scholar at OGI Portland (USA) for three months during the summer of 2001. He has done his post-doctoral studies (PDF) at Carnegie Mellon University (CMU) Pittsburgh (PA, USA) during April 2007 to July 2008. He is an author of about 185 papers published in national as well as international journals, conferences, and edited volumes. He is a life member of the CSI, IE, IUPRAI, ASI, IETE, ORSI, and ISTE. He has reviewed papers for reputed journals and conferences. Dr. Suryakanth V Gangashetty has about 24 years of experience working with speech processing technologies. He has participated in Speaker Identification, Speaker Verification, Anti Spoofing challenge and Text to speech synthesis benchmarks. He has worked on Speech Signal Processing for a Large Vocabulary Continuous Speech Recognition, while at CMU Pittsburgh (USA). He has been part of ASR and TTS projects funded by MeitY. He has undertaken many projects sponsored by the Indian Government (such as MHRD, TDIL) and the IT Industry (such as Samsung). He has also executed International Collaborative projects (UKIERI) in the area of speech processing. He has guided and currently guiding PhD scholars in the areas of Speech recognition, Speech synthesis, Voice conversion, Language identification. Speaker recognition, Speech enhancement, Audio scene classification. Code mixed speech processing, Dialect identification. Emotion recognition and prosody processing. He has experience in the development of speech to speech translation system for the languages English, Hindi, and Telugu. His research interests include Speech Pro
cessing, Neural Networks, Machine Learning, Natural Language Processing, Artificial Intelligence. He was local Organizing Chair for the INTERSPEECH-2018 conference which happened in India in September 2018 held at Hyderabad. Currently he is one of the member of the ”NATIONAL LANGUAGE TRANSLATION MISSION (NLTM)Consortium project Speech Technologies in Indian Languages (Speech quality control). He is also one of the Co-PI of the ICMR project. He is also Principal investigator of project Sponsored by Naval Research Board (DRDO) New Delhi. India.

ISSN 2210-142X

Open Research Issues and Tools for Visualization and Big Data Analytics

Rania Mkhinini Gahar1,2, Olfa Arfaoui1,3 and Minyar Sassi Hidri4
1National Engineering School of Tunis, University of Tunis El Manar, Tunisia
2OASIS Research Laboratory
3RISC Research Laboratory
4Computer Department, Deanship of Preparatory Year and Supporting Studies, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

ABSTRACT

The new age of digital growth has marked all fields. This technological evolution has impacted data flows which have witnessed a rapid expansion over the last decade that makes the data traditional processing unable to catch up with the rapid flow of massive data. In this context, the implementation of a big data analytics system becomes crucial to make big data more relevant and valuable. Therefore, with these new opportunities appear new issues of processing very high data volumes requiring companies to look for big data-specialized solutions. These solutions are based on techniques to process these masses of information to facilitate decision-making. Among these solutions, we find data visualization which makes big data more intelligible allowing accurate illustrations that have become accessible to all. This paper examines the big data visualization project based on its characteristics, benefits, challenges and issues. The project, also, resulted in the provision of tools surging for beginners as well as well as experienced users.

Keywords: Big Data, DataViz, Cloud, Big Data Analytics, Business Intelligence.

1. Introduction

Every day we generate 2.5 trillion bytes of data. In fact, 90% of the world's data was created in the past two recent years only [1]. The data comes from different sources such as the sensors used to collect climate information, social Media Posts, digital images, and online videos, online Purchase, transnational records and GPS (Global Positioning System) signals from mobile phones. This data, which has resulted essentially from the meeting of three elements which are Internet, social networks, and smart devices (computers, tablets, smartphones, connected objects...), is called big data or massive data. It is considered very interesting according to the pertinent information that may contain. Actually, we note that 13 of business leaders' decisions are based on information they don't trust or don't have and half of them say they don't have access to the information they need to do their job and 83% of CIOs (Chief Information Officer) cite analytical business intelligence (BI) as part of their plans to improve their competitiveness.

Moreover, 60% of CEOs (Chief Executive Officers) need to improve the capture and understanding of information in order to make decisions more faster [2]. For example, data could be analyzed to i) detect customer feelings and reactions or critical or life-threatening conditions in hospitals to intervene in time; ii) predict weather patterns to plan the optimal use of wind turbines and make decisions based on transactional data in real-time; iii) identify criminals threats from videos, sounds and data streams; iv) studying student reactions during a lesson and predict which ones will succeed in the basis of statistics

and models gathered over the years (big data in the education field).
Research on big data analytics is entering a new phase called fast data where multiple gigabytes of data arrive in the big data systems every second [3], [4], [5], [6], [7], [8]. Modern big data systems collect inherently complex data streams due to the 3 basic Vs which are Volume, Velocity, and Variety and to which are added Veracity, Validity, Vulnerability, Volatility, Visualization which consequently give rise to the 10Vs [9] of big data. The well-designed big data systems must be able to deal with all 10Vs effectively by creating a balance between data processing objectives and cost (i.e., computational, financial, and programming efforts). Data collection and storage capabilities have allowed researchers in diverse domains to collect and observe a huge amount of data. However, large data sets present substantial challenges to existing data analysis tools [10].

We will focus in our paper on one of the most important big data' Vs which is data visualization. One of the most crucial and useful tools for comprehending corporate information is data visualization. Because of this, a picture truly is worth a thousand words. Data has been visually represented by humans for hundreds of years. We've collected data, organized it, and presented it in maps, charts, and graphs to tell a richer and deeper story than it may have otherwise. The data boom coincides with the technological boom. Additionally, we have been able to process ever-increasing volumes of data quickly thanks to the same technology. Although they might not be immediately apparent in the first text format, trends, patterns, and other insights are quickly identified utilizing data visualization software.

The most effective strategy changes to visual data displays after reports and dashboards take their place since they can fit a lot of information into a little amount of space. Examining the extensive data sets and graphic presentations that enable quick and accurate translation might take hours, days, or even weeks. Thanks to advanced technology, many data visualization tools allow for interactive functions. This f lexibility provides the ability to switch and change quickly, which helps the user to discover and learn about alternative viewpoints. This comprehensive, interactive presentation can rarely be achieved quickly by processing raw data without visualization software. The information industry frequently faces the difficulty of the quantitative component.

Knowing that decisions are made as a result of visual representations requires a solid comprehension of data. In the absence of context, visuals are ineffective. But there is an easy target is just to let the workers and the tools perform their jobs. As long as you utilize the appropriate tools and the individuals conducting the data analysis are aware of where the data originated from, who can use it, and how it will be used. The data visualization will next be translated, processed, and on a more clearer course for making those important decisions. Data visualization's significance in the world of corporate information is being realized more and more every day. They have the ability to not only supply useful data but also understand how to process it, which guarantees the organization stays competitive. This is because they are high-performance analytics tools that offer better ways to analyze data faster than ever before.

Visualization is critical in today's world. Big data is difficult to visualize. Due to in-memory technology limitations and low scalability (scaling up), functionalities and development time response, visualization tools, and current big data are faced with technical challenges. Traditional graphs cannot be relied upon to attempt to plot a billion data points. We, therefore, need different ways of representing data. If we take into account the multitude of variables resulting from the variety and speed of big data and the complex relationships that unite, we can see that developing a visualization significant is not so

so easy.
Spreadsheets and reports stuffed full of numbers and algorithms are far less successful at communicating meaning than reports and charts that show enormous amounts of complex data. Because of the constraints of in-memory technology and their inadequate scalability, functionality, and response time, current large data visualization solutions suffer technological difficulties. When plotting a billion data points, we can't rely on typical graphs, thus we need alternative methods, such as data clustering or the use of treemaps, sunbursts, parallel coordinates, circular network diagrams, or cone trees. [11].

The rest of this paper is organized as follows: The DataViz' presentations are presented in section 2. Section 3 presents its benefits. Section 4 introduces the characteristics of such DataViz project. Section 5 discusses some DataViz tools surging for beginners as well as well as experienced users. Section 6 revealed the main data visualization challenges. The DataViz issues are highlighting in section 7. Overall discussion with limitations are stated in section 8.

2. The DataViz and you: presentations

Many data scientists define data visualization in different ways. Indeed, they agree that it is indeed a visual form to visualize by facilitating access to it. Another data experts' group agrees that data visualization is meaningless if it does not encompass understanding, exploitation and decisionmaking, speed, and information sharing.

A. DataViz features

The value of data visualization lies in its ability to meet three main imperatives namely interpretable, relevant and innovative.

• Be interpretable: In a context where the volume of data is exploding with the exponential growth in the use of the Internet and in particular Google, so-called unstructured data experiences the same evolution. But a data visualization starting from these data which would not be interpretable, that is to say clear, would be useless. There must be some clarity regardless of the volume or source of the data [12].
• Berelevant: At a time when big data is a central issue for companies, several techniques to process this mass of information in a relevant way must be put in place. Relevance is linked to interpretability. The data visualization must make it possible to answer questions in a defined context and aimed at specific objectives. Data sources must be reliable. Data integrity is the basis for meaningful data visualization. You must ensure that your information is correct and up to date. It is necessary to sort the data for optimal analysis and to consider using all the data at your disposal. This allows you to cross-reference information and thus bring out a more complete analysis to better support your digital marketing department[13].

• Be innovative: Finally, data visualization is only of interest if it brings new information, and originality, if it gives a perspective unpublished on a subject. It provides a different perspective, to illustrate an analysis [14].

B. Brief history of DataViz

Originally, it was a simple human limitation that spawned the Data Visualization approach: our brain is simply unable to easily process large volumes of raw data to extract useful information. Maybe we can do it occasionally, but certainly not every time, let alone many times every day. However, Data Visualization is not just about the graphical representation of data. It is also a story that is told with the

help of these representations.

The first successful expression of this approach is a flow map that tells the story of the Russian campaign led by Napoleon. We owe it to the engineer Charles Joseph Minard who represents, at the beginning of the 19th century, the story of the colossal human losses of the Russian campaign during which the Napoleonic army arrived in Moscow with less than a quarter of its starting squad.

Shortly after, nurse Florence Nightingale had the idea of using graphic representation to allow her reader to compare facts with complex correlations [14].

She presented, for the attention of Queen Victoria, the main causes of death of British soldiers engaged in the Crimean War. His graphic support, dated 1858, allowed him to eloquently highlight that epidemics were much more devastating on the workforce than the injuries suffered in combat. Florence Nightingale has therefore used a graphic support of data, presented in figure 1, to communicate information, of course, but also to convince, that is to say, to orient the conclusion that one draws from it.

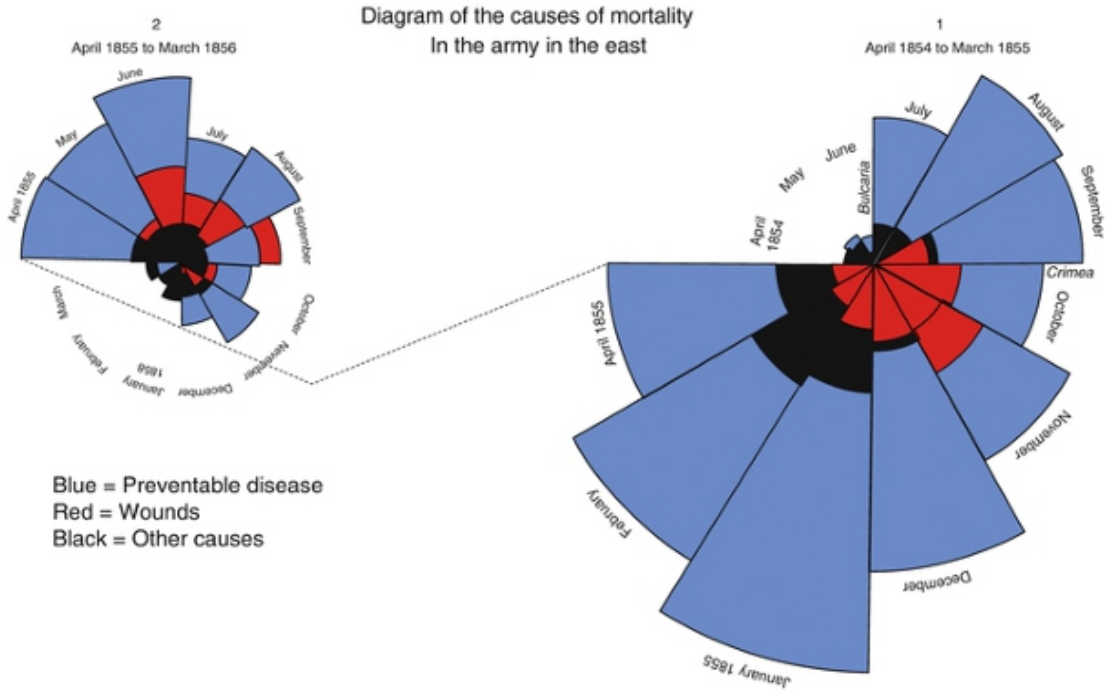


Figure 1. Florence Nightingale's Diagram [15].

Until the 19th century, Data Visualization therefore evolved through isolated, independent attempts, which explored both the possible fields of application and the possible graphical approaches. In other words, the discipline was forging a vocabulary but it still lacked common rules– that is to say, a grammar.

This grammar was laid down by Jacques Bertin who in 1967 developed the real bases of graphic language.

In figure 2, Jacques Bertin defined graphic semiology, i.e. the elements that can be modified in a Data Visualization to represent information. Identifying and clearly defining these graphic variables (color, size, surface) was simply the grammar that was missing from the graphic language.

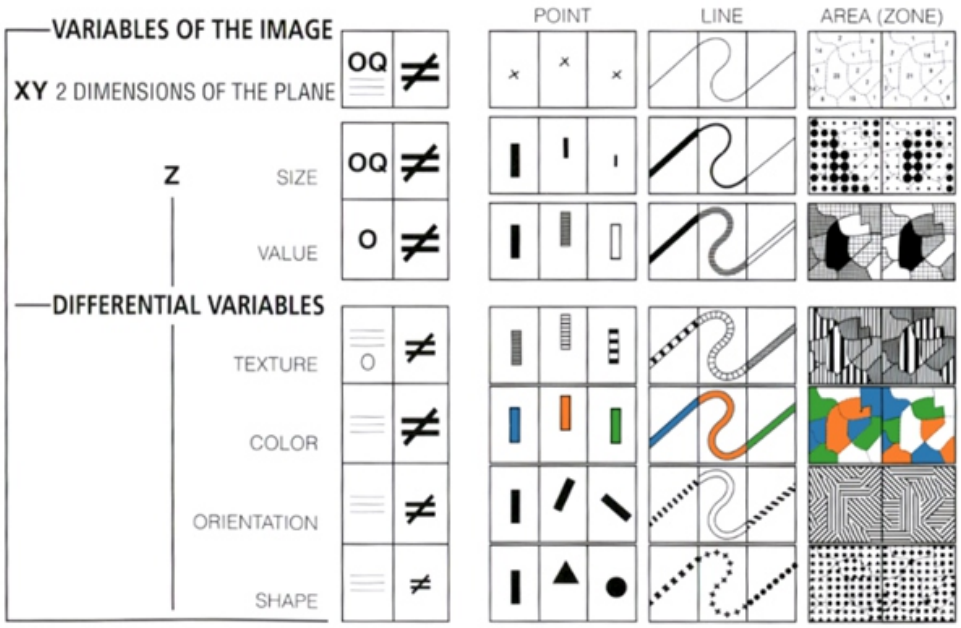


Figure 2. Jacques Bertin's graphic semiology [16].

It is by applying these new– but common– rules that we have gradually managed to remove superfluous graphic elements and define what characterizes a relevant data visualization.

C. The DataViz is pretty, but what is it for?

Data visualization concerns a wide range of business sectors. If we stick to our panel, several areas are represented, namely consumer goods, business services, industry, media, marketing and advertising, scientific research, public service, telecom, transport, logistics, etc... All companies use data visualization either to do certain things they were already doing better (optimization), or to enrich their activity with value-added tasks (innovation). In terms of optimization, DataViz allows, for example, to Lagard`ere Active to accelerate the production of its reports and STMicroelectronics to make its manufacturing process more efficient;

In the area of innovation, PagesJaunes discovers and rectifies, thanks to data visualization, the shortcomings of its indexing, Voyages-sncf.com launches a new innovative service (Mytripset), Alcatel Lucent imagines the mobile applications of tomorrow, etc.

D. Dual purpose of optimization/innovation

More specifically, 5 use cases illustrating this dual purpose optimization and innovation: the use of DataViz to:

• Improve the company management: One of the first uses of data visualization is to contribute to more effective management of activity and performance, oriented towards action.

In our modern economy, all directions need to manage their organizations as finely as possible by relying on numerous and rich data and by producing relevant dashboards. "A picture is worth a thousand words" and nowadays "the image takes precedence often on words" [17].

Data visualization is the art of telling figures in creative and gameful way [18].

• Improve the customer relationship: Marketing and customer relationship management are two functions of choice for data visualization. We visualize customer data to improve Customer relationship management (CRM) multi-channel. In addition, DataViz is better to qualify the customer base.

• Define the company's offer: At the opposite end of the spectrum, data visualization provides companies with tools to better define their offers. Moreover, the exploration of collected customer data and the ability to test different hypotheses prove particularly valuable.

• Contribute directly to the company business: The DataViz serves to better understand your competitive positioning. That's why Data visualization could be considered a relevant source of value and differentiation which immediately identify customer postponements. Here, data visualization could directly influence the business economics model [19].

• Empowering citizens: Data visualization also finds its usefulness outside the walls of the company. It can contribute to better informing the citizen and therefore giving him the means to act. In fact, company can allow citizens to think through Citizen DataViz projects may take a longer turn activist [20], [21]. This is the case, for example, of the pariteur of France Televisions.

E. DataViz and Infographics

The Dataviz takes the information a person needs and presents it in an easily understandable way. Infographics are a mix of Dataviz, journalism, and marketing. They use strategically chosen data visualizations and lexicons to explain a complex story easily. The confusion in terminology is understandable however the terms are not interchangeable. Both turn data into easy-to-understand visualizations. These tools are extremely powerful when it comes to explaining numbers in an educational way to people who are reluctant to analyze data. This is their only common point. Here is a definition for both forms of presentations.

Some distinction points between these two terms can take place in Table I.

F. Main reasons for using DataViz

Three main reasons explain the use of data visualization namely confirm or refuse hypotheses on a market, educate and explore.

• Confirm or refuse hypotheses on a market: The DataViz can then take the form of a dashboard, making it possible to decide while having a global vision of the studied market.

• Educate: Internally, companies use DataViz for research work reporting or brainstorming sessions. It can be a good complement to creative approaches such as gamification.

• Explore: This is the most futuristic aspect of data visualization, which certainly will develop. Dataviz can help build predictive models. We are then in the field of data analysis

G. DataViz: buzzword or real innovation?

The data that are thus available to professionals to guide them in their decisions are increasingly numerous and multistructured. But how not to be overwhelmed and make it a real tool for reflection and decision-making? How to obtain answers to fundamental questions whose answers are for the moment unknown?

It is in the face of these requirements that DataViz takes on its full meaning. The representation of data in the form of images makes it easier to understand them. There are several definitions of DataViz in the academic and industrial state of the art. These different definitions all converge on the fact that DataViz is a way to give meaning to data in order to extract information from it and therefore exploit it. Dataviz not only enables intellectual understanding but also transforms a set of raw data into actionable information.

In addition, it accelerates the understanding, decision, and action that we have just mentioned. It is also a mode of communication, allowing data not to remain confined to the world of BI or statistics but to infuse the entire organization and become a support for decision-making and collaborative work.

DataViz has many uses and leads to a variety of benefits for the organization. First, it contributes to more effective management of activity and performance, oriented towards action. This improvement in management is manifested by taking a step back in addition to other tools whose horizon is in the shorter term. In other words, DataViz can be used as a decision-making, and strategic tool, usable by a local manager to manage his performance.

Another important use of DataViz is the reinvention of customer service to improve its efficiency. To improve the management and understanding of their KPIs (Key Performance Indicators), SFR uses DataViz to identify causal relationships in their data sources in order to find hidden

TABLE I. Infographics versus DataViz

Infographics	DataViz
Promotes the information understanding that we already know by representing it in graphic form.	Bring out information that was unknown by analyzing data presented in graphical form.
Modest data amount.	Huge data volume.
Good design makes a product useful.	Good DataViz makes information useful.
Good design is aesthetic.	Good Data is aesthetic.
Good design makes a product understandable.	Good DataViz makes information understandable.
Good design is honest.	Good DataViz is honest.
A didactic approach focused on others.	Self-knowledge tool.
Understanding support.	Decision support.

patterns through their main sales channels. A good customer relationship is based on perfect knowledge of the customer himself. What are their characteristics and behaviors? How to segment and classify them? The exploration capabilities in the data enabled by data visualization find their full meaning in providing answers to these questions.

Another key point about DataViz is its ability to foster innovation and its potential to get the business to consider new possibilities. In particular, it is a testing ground for new modes of interaction with users.

3. DataViz's benefits

In a context of ever-increasing and often highly complex volumes of data, DataViz has many advantages. Data visual
ization is far from being an accessory intended to embellish your website or your presentations. Synthetically, we can say that DataViz improves the data understanding, the data communication, the decision-making, and the ability to innovate.

A. Dataviz makes it easier to understand data

With big data advent and the proliferation of data sources, companies are increasingly using data visualization. These visual representations make it easier to understand raw data and thus help in decision-making. Big data is not merely more data; it is data that is so vast, so varied, and collecting so quickly that typical procedures and methodologies, including "normal" software like Excel, Crystal reports, or other programs, are ineffective. DataViz makes it possible to make the most comprehensible data important and what they mean, regardless of the audience concerned. Its effectiveness is based on the fact that a majority of us grasp and retain information better when it is represented visually. The

following image illustrates this fact, which was studied by an American psychologist.

Unlike a table filled with figures, DataViz helps to highlight information that seems complex or drowned in a large number of parameters. The following example illustrates this fact well. For example, we want to analyze the life expectancy by country. Table II presents an extract of values from the top 5 countries since the values file is very large.

But is it ideal for ordering easily and between countries quickly? And explain it in a few seconds to his audience? If we transcribe this information on a map, with more or

TABLE II. Life expectancy per country.

	Country	Life expectancy	ISO-code
0	Afghanistan	64.5	AFG
1	Algeria	76.7	DZA
2	Andorra	81.8	AND
3	Angola	60.8	AGO
4	Antigua and Barbuda	76.9	ATG

less bright colors depending on the strength of the index, everything becomes clearer. We want to quickly understand which countries have the highest rate of life expectancy.

Figure 3 allows us to assess at a glance countries with the highest rate of life expectancy. Analyzing an Excel file is much faster by visualization, especially if the data is large and complex.

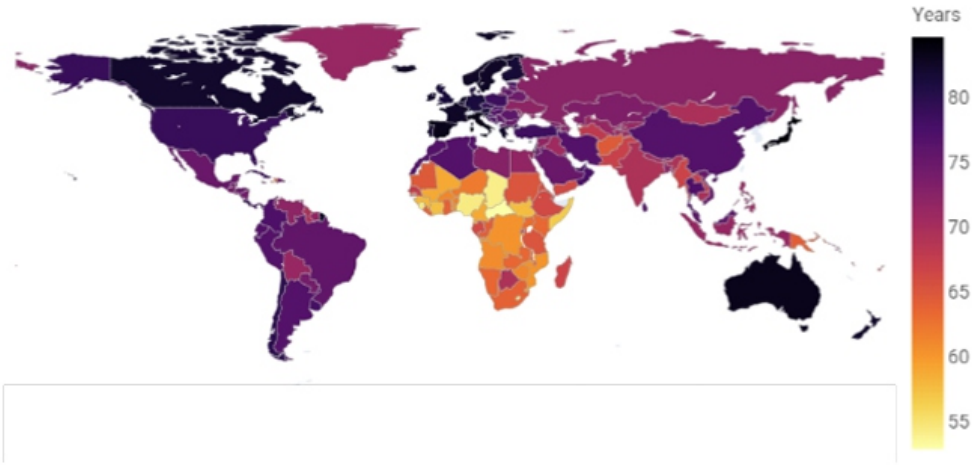


Figure 3. Life expectation by country (Source: Kaggle / WORLD

DATA by country (2020).

Our brain needs less than 250 milliseconds to enter (1), understand (2), and respond (3) to information under visual form. Whereas comparing raw data in tabular form requires an effort of memory that quickly reaches its limits. Ultimately, data visualization invites us to take up the classic distinction between data, information, and knowledge. If the data are unitary, raw elements, reflecting reality, the information is their coherence to give them meaning.

B. Dataviz improves communication

Data not only reflect reality, but they also are not just a steering lever. They are also communication tools. Unfortunately, few of us are fluent in our language. Most of us need an interpreter to make us penetrate the intelligence of the data.

This is where data visualization comes in. Moreover, Communication is the major asset of the DataViz for coordinate the teams. Indeed, the infinite volumes of data stored by companies are generally only within the reach of data analysts and other technical profiles. But once put into images, this data is accessible to everyone without any prior training. The support generated by DataViz makes it possible to unify the discourse and convey unambiguous messages. We can truly speak of democratization of access to given [22].

C. Dataviz optimizes and accelerates decision-making

This is the logical continuation of an easier understanding of the data. DataViz allows the development of interactive dashboards. Unlike static charts, such as those in Excel, interactive DataViz allows the exploration of data in depth, with less effort. In a few clicks, it is now possible to release correlations between the operational actions put in place, the performance, and their impact. It is also much easier to compare its indicators to the market, and competitors. Once the data has been clarified and better identified, decision-makers can focus on the essentials and make choices in a way more simple.

D. Dataviz promotes innovation

Finally, we cannot close this part on the benefits of data visualization without examining its potential in terms of innovation. Indeed, DataViz is also a field of research that can encourage the company to consider new possibilities. In particular, it is a testing ground for new ways of interacting with users [23].

4. Characteristics of DataViz projects

The deployment of DataViz software is a crucial step that must be perfectly orchestrated to guarantee the success of the project. This later must be both fast and light [24].

A. DataViz project's speed

The first characteristic of DataViz projects is that they are, all in all, quick to conduct. Their big advantage is in particular to reduce the time between the launch of the project and the ability to show a first operational version [21].

There are several reasons for this speed. One of them is that DataViz projects require hardware and software resources that interfere little with existing architectures. They, therefore, do not require long and complex budget validation cycles: the necessary environment can be available in a short time. Another reason is that data visualization lends itself well to POC (Proof Of Concept) type approaches experimentation, trial and error iteration loops. Dataviz projects have a very empirical side, completely in phase with current development approaches, such as agile methodology, rapid Agile development, or Scrum.

B. DataViz project's lightness

Lightness is involved in the technical resources required which are quite light. Example: Recent technological developments, such as the development of JSON-type formats (DS.JS3), put us in a direct connection with data. Thanks to these formats, we can recover varied data from all horizons, using standard applications. A real human-data interface is thus being established.

C. Success key factors

To succeed in a data visualization project, it is necessary to bring together key success factors that can be classified into three categories. First of all, there are the classic good practices of any project: ensuring preparation and planning, choosing the right scope, implementing the appropriate methodologies, etc. The second category is that the DataViz project concerns data: their targeting, quality, respect for confidentiality, and access authorizations, are therefore essential. Finally, of course, ergonomics and graphic intelligence play a key role in the acceptance of DataViz and its effective use (although, by the way, these aspects should be part of any IT project, etc.).

1) Prepare the project well

The first secret to success is preparation. This is understood on two levels: the content of the data visualization, on the one hand, and the project approach, on the other hand. DataViz is not a panacea and, to be useful, it must be well thought out. On the methodological aspect, the preparation consists of implementing establishes a process for collecting, analyzing, and representing data that is viable over time, but also sufficiently flexible.

2) Target data and visualizations according to the profile of the users

A second success factor is related to the nature of the data visualization, namely a communication tool. However, for communicating effectively with someone, it is important to meet his need, with clarity and in a form he understands. Three key questions to ask yourself to choose the best representation:

• What question do we want to answer? All the graphs do not make it possible to present the same analysis (distribution, evolution, decomposition...). Hence the importance for the designer to question his intention.

• Whoare we talking to? Is he an expert or a layman? What should he do with the information (e.g. retain the information for later or make an immediate decision)?

• In what context is the interlocutor? The good reception of the graphic does not only depend on the graphic itself, but also the intellectual and visual availability of the reader. All that the graphic designer can do is try to anticipate this greater or lesser availability, in order to choose the most suitable representation.

The difficulty increases when the data visualization must address different audiences. It is then necessary to provide modes of representation adapted to each of them. This is the case, for example, of the Belgian FPS Economy, which communicates with both the general public and professionals.

3) Start on small perimeters, to learn

Another good practice is to ”get your hands dirty” on first restricted perimeters, if possible as controlled as possible. This allows you to move forward, without too many risks, in trial and error mode until a first satisfactory solution is obtained.

4) Ensure data quality at source

In BI, you reap what you sow or, to put it more lapidary way: garbage in, garbage out. In other words, if we want data visualization to be able to communicate the right messages, make it possible to make informed decisions, to explore unknown territories, there is one condition to be met above all: to have quality data entrance. In return, DataViz improves the quality of the data. First, because it compels a certain discipline; then, because it also visualizes... the non-quality of the data. Repeated outliers may appear at first glance as oddly placed dots, for example.

5) Focus on cooperation between several departments

Another ingredient of success lies in the cooperation between the actors of the project. DataViz thus contributes to breaking down the silos that may exist in the company and contributes to greater cross-functionality.

6) Train the teams

Data visualization does not require side training users. On the contrary, we can say that it is fully successful when it is immediately adopted. For this, simplicity and intuitiveness are essential. But offering a simple rendering can be extremely complicated. This is why the training of those who produce the visualizations is an undeniable plus.

7) Using aesthetics as a lever for appropriating information

Data visualization cannot be reduced to a representation aesthetics of data. One can make pretty representations that are perfectly useless. But that's not to say that aesthetics don't play a role. Used well, it is an essential criterion of efficiency for DataViz. In this concern to combine aesthetics and efficiency, companies have every interest in being imaginative and going beyond traditional Excel-type charts, if that makes sense.

D. Pitfalls to avoid

If certain good practices maximize the chances of succeeding in your data visualization project, you can expect, as with any project, to encounter difficulties such as:

- The risk of overloading it with information.
- A general management that does not necessarily perceive an interest of data visualization immediately.
- Skepticism about the performance of the tool.
- Apitfall to avoid: forget your classics: We should not confuse simplicity with simplism [25].

E. The DataViz's impact on the relationship between IT and business lines

Data visualization projects have the particularity, as we saw previously, of offering great autonomy to users. users. This is why, in this area, the relationship between IT and the Professions are set to evolve. It has happened that friction has arisen, with the IT Department feeling deprived of some of its prerogatives. These tensions have unfortunately been maintained by some providers of DataViz solutions by addressing only the Business Lines without going through the IT Department to win contracts [26].

Thus, we cannot speak of a loss of prerogatives of the DSI (Dimensional Strategies Inc.) or the BI teams. Simply, data visualization raises new questions about how to represent data, about the distribution of roles and responsibilities between the business lines and the IT department, and about how to conduct projects [27].

CIOs understand this. Even those who were initially reluctant are realizing that data visualization is not a threat, and are softening their stance [28]. In truth, data visualization is a chance for CIOs and BI teams. On the one hand, it will relieve them of time-consuming tasks, allowing them to focus on their missions with higher added value. On the other hand, they even have a unique opportunity to invent a new form of BI and relationship with business. What we can remember is that data visualization, even if it provides great autonomy to the business lines, is a question that should interest the IT department and the BI teams, quite simply because it touches the data. Autonomy of businesses is useful if it is implemented smartly and if it allows them to obtain even more value from the CIO and the BI. Presumably, by accustoming the Professions to speak the language of data, thanks to graphics, the DataViz will

contribute to the taking of awareness of the value of data. It can therefore play a role unifier, at the service of business creation.

5. Which tools for which data visualization?

Information visualization faces increasing hurdles as the big data era progresses. First of all, the amount of data that needs to be visualized surpasses the size of the screen.

Second, a typical computer cannot be used to store and process the data. A big data visualization solution needs to offer perceptual and performance scalability to solve both of these issues.

In this section, we will focus on some data visualization tools for beginners as well as for experienced users.

A. Tools accessible to beginners

Tools for beginners are available to allow them to create DataViz without resorting to programming or its basis and no expertise is required.

1) Office software and extensions

a) Excel

Excel remains one of the basic tools for data visualization [29], [30]. The maximum number of values in a column is about 1,999,999,997 [31]. Third parties create Excel add-ins to provide Excel users with extended functionality and save their time and effort. Developing these add-ins requires coding expertise in languages such as XML (eXtensible Markup Language) and VBA (Visual Basic for Applications) and providing an easy-to-use interface that complements Excel.

Table III overviews some Excel add-ins for DataViz.

Table III overviews some Excel add-ins for DataViz.

b) LibreOffice

LibreOffice is a free and open-source office suite, derived from the OpenOffice.org project, created and managed by The Document Foundation. Extensions are provided to enable various activities, including visualization. LibreOff ice extensions are software add-ons that you can install in addition to the core LibreOffice apps to extend the capability of the suite in one or all of the programs (Writer, Calc, Impress, etc.).

Some examples of LibreOffice extensions dedicated to data visualization can be presented in the table IV.

2) Online office suites

Whenchoosing between data visualization tools, one option worth considering is Google Sheets. Google's spreadsheet application can be used to generate charts, tables, and even maps that can be embedded in a website. They're easy to make and can be configured to update automatically [32].

It's not for every visualization need. There are some tasks that need for more intricate data visualization strategies and customisation than what Google Sheets can offer. Google drive extensions can also help novice users to perform data visualizations, namely Fusion Tables [33], Slemma [34], Geckoboard [35], VizyDrop [36], BIME Analytics [37], Cyfe [38], and Datahero [39], etc.

3) Office 365

Microsoft 365 is made up of the Office suite (Word, Excel, PowerPoint, Outlook, OneNote, Publisher, and Access), as well as a set of online services such as OneDrive, Exchange Online, SharePoint Online,

Teams, and Yammer.

The Officesuite allows work in offline mode like a perpetual suite, which distinguishes it from Office Online, which is used from a Web browser. The principle of Microsoft 365 is to be updated as new versions of Office are released [40]. It provides also some integration apps to visualize your data in an interpretable, innovative and relevant way. Table V describes someones highlighting their main functionalities.

4) Simple online tools

a) Tableau

Tableau Public is a free platform for exploring, creating, and publicly sharing data visualizations online [41].

b) Canva

Unlike other charting tools, Canva is quick and easy. There's no learning curve– you'll have a beautiful graph or chart in minutes, turning raw data into something visual and easy to understand [42].

c) Plot.ly

Plotly is an open-source visualization library for data visualization and analysis. It provides many products including Dash, Chart Studio,, a Python framework, R, and recently JULIA for building fast, easy and powerful analytical applications. It gives the hand to draw several types of the graph such as 3D graphs, histograms..., easy to use and handle, totally free, and very interactive and flexible.

d) PowerBI

Is a data analysis solution from Microsoft. It allows for the creation of personalized and interactive data visualizations with a simple interface enough for end users to create their reports and dashboards [43].

e) Chartblocks

This is an online charting software. It helps to create basic charts quite quickly and to import more data from different external sources [44].

f) Periscope Data

Is an effective platform for data analysis. It may compile all of the company's data and produce reports. With this tool, we can quickly transform our data into a report or graph that is simple to interpret. [45]. For data consumers who frequently need data, Periscope Data enables analysts to transform their SQL searches into interactive dashboards, charts, and reports. The groundbreaking data warehouse technology from Periscope Data links to your databases in a flash to provide extraordinarily speedy, low-cost query processing. Workflow hurdles are removed, and data literacy is promoted throughout your organization, thanks to unlimited users and no query limits. [46].

g) Holistics

Is an intelligent data reporting and business intelligence platform that enables us to resolve our data-related inquiries and problems without the need for technical support. For business and data teams, it eliminates the aggravation of request queues. Holistics allows business users to access their data without writing SQL (Structured Query Language) or interfering with data teams. Data teams can create and manage a set of business KPIs using Holistics [47].

h) Cluvio

It is a cloud-based analytics and BI tool that enables businesses to use a dashboard to examine data. The solution enables query execution, results filtering, and data visualization on charts and diagrams. An SQL editor, adaptable

TABLE III. Some Excel add-ins for DataViz.

Add-in	Description
Filled Map	Used to display high-level chart data within a map.
3D-Mapping	A three-dimensional (3D) data visualization tool called Microsoft 3D Maps for Excel enables you to see the data in novel ways.
Bing Maps	With the Bing Maps add-in, Excel users can quickly plot locations and display their data using Bing Maps.
Radial Bar Chart	It displays data from standard bar charts on a circular pattern. With this representation, visual analysis gains the advantages of both bar charts and circle graphs. The Radial Bar Chart has a distinctive style that makes it a very adaptable representation.
XLMiner Data Visualization App	It instantly visualizes data in your Excel spreadsheet and lets you adjust the variables plotted on each axis, zoom in and out, add filters to highlight important data, and use textit size by and textit color by categorical variables for quick insights.

TABLE IV. LibreOffice Extensions.

Extension	Description
GeoMap	Inserts map images directly into your document from address information.
GeOOo	His free solution suggests using the LibreOffice tools to create thematic maps.
ClusterRows	A Is a LibreOffice Calc extension that groups rows into clusters and colors the clusters to show them.
OOoHG Gallery	Includes a *Gallery* that contains 1600 maps, diagrams, and graphs for geography and history organized into 96 categories (in bitmap and vector graphic format).
OpenStreetMap Presentation	It uses penStreetMap Datas as background for nice presentations.

TABLE V. Data Visualization Apps Integrated with Microsoft 365.

Apps	Description
Workday Adaptive Planning	Is a cloud-based enterprise planning platform that provides modeling, budgeting, and forecasting capabilities to enable firms collaborate on planning. It offers capabilities like workforce modeling, balance sheets, and spending management, among others.
Coras	It is an enterprise decision management platform suitable for IT, government, legal, and marketing teams. It can be adapted to any type of business and easily mapped to existing processes and governance.
Microsoft Power BI	It enables the creation of personalized and interactive data visualizations with an interface simple enough for end users to create their reports and dashboards.
Cyfe	Users can monitor and analyze data spread across all of your web services, including Google Analytics, Salesforce, Google Ads, MailChimp, Facebook, Twitter, and more, from a single spot in real-time with this all-in-one dashboard software.
DBxtra	It is an ad-hoc reporting and business intelligence system that gives companies the tools to create and distribute unique reports on key performance indicators. It has capabilities including an Excel reporting service, a report and dashboard designer, online report distribution, and a report scheduler.
MicroStrategy Analytics	It equips businesspeople with self-service tools to study data and share insights in minutes, enabling them to make quicker, smarter business decisions.
Kepion	It is a Microsoft BI-powered cloud-based business planning tool. In a single, centralized platform, it combines budgeting, forecasting, BI reporting, and intuitive modeling technologies, empowering users to create and schedule applications that are tailored to the operations of their firm.

R scripts, push notifications, customer dashboard sharing, and more are some of its primary features [48].

I) Klipfolio

is a cloud-based tool for building and sharing real-time dashboards for use on mobile, TV, and online browsers. [49].

j) Clicdata

ClicData is a 100% cloud platform. It offers a more modern vision of the software by offering in particular the possibility of importing data regardless of their format and cross-referencing information

from different tools [50].

k) Qlik Sense

An effective visualization presents the relationships between many values and allows you to analyze data at a glance. Qlik Sense offers a wide range of visualizations and charts. Each chart excels at visualizing data in various ways for different purposes. Charts should be selected based on the data you want to see in them [51].

l) Chartio

With the help of the cloud-based BI and analytics tool Chartio, users can quickly evaluate data from business apps and visualize it using a variety of customisable charts. Due to its straightforward Interactive and SQL modes, Chartio is appropriate for both professionals and those with no prior technical knowledge [52].

m) DataWrapper

Another excellent tool for data visualization is Datawrapper. With DataWrapper, you can simply generate charts, tables, and maps that are readable on any device. As a non-commercial platform, Datawrapper is best suited for schools and small businesses that require simple data visualization tools [53].

n) Venngage

Venngage is a web application for creating a range of data visualizations including infographics, posters, reports, and promotions [54].

o) Piktochart

Piktochart is a web-based graphic design tool and infographic maker. We can create bar charts, maps, line graphs, scatter plots, and more [55]. It can also synchronize with Google Spreadsheet or SurveyMonkey to retrieve data and thus create interactive graphs or tables.

p) Infogram

Infogram is a web-based data visualization and infographics platform. It operates as a data visualization tool to make data easy to understand, discover unknown facts/outliers/trends, visualize relationship patterns, and ask better questions.

q) Raw

RAWGraphs is an open-source data visualization platform designed to make it simple for anyone to visualize complex data. It tries to fill the gap between vector graphics editors like Adobe Illustrator, Inkscape, and Sketch and spreadsheet programs like Microsoft Excel, Apple Numbers, and OpenRefine [56].

r) Wordle

Wordle is a tool for altering word clouds [57] that include participation from the Las Vegas 2009 Olympics. The fundamental advantage of Wordle is that it enables neighborhood-preserving editing, which retains words in predictable and nearby areas both during and after editing. An illustrative example made with wordle for the DataViz context can be presented in figure 4.

Figure 4. DataViz with Wordle.

s) Easel.ly

People may simply visualize information on Easel.ly by quickly creating infographics and data visualizations. No prior design experience is necessary. A tool for designing infographics called Easelly can turn any visualized content into any kind of information. It offers a variety of templates, themes, and objects so users may alter specific details in their projects [58].

B. Tools accessible to experienced users

1) JavaScript libraries

JavaScript libraries and frameworks facilitate the development of websites and applications with a wide range of features and functionality– all thanks to the dynamic, flexible and attractive characteristics of JavaScript. According to a 2020 StackOverflow survey, JavaScript remains the most widely used programming language (for the 8th grade), with 67.7% of respondents using it. Table VI including the top 15 libraries dedicated to visualization will be able to describe them better.

2) Dashboard builders

a) Google Data Studio

It is an online application for data visualization that assists users in transforming data into educative reports and engaging dashboards. By producing engaging reports like this one, it is a powerful tool that enables marketers and business owners to use their data efficiently [59].

In essence, Google Data Studio is a condensed version of software for data visualization, such as Tableau and Clickview. Data Studio is not a data source, in contrast to tools like Google Analytics or HubSpot. It gathers data from several sources, does analysis on it, and then enables you to produce interactive reports, charts, and dashboards rather than collecting the data.

b) Toucan Toco

Is a cloud-based data visualization tool. Intended for non-technical business executives, the objective of this highly configurable data visualization solution is to provide essential information and data for decision-making [60]. Toucan Toco also develops APIs that allow it to integrate with other IT solutions, such as Cognos Analytics and Salesforce. To retrieve the data to be used and then displayed, this advanced reporting tool is thus able to connect to more than a hundred applications: Excel, Google

Analytics, and Microsoft SQL Server.

c) Data Hero

DataHero is a cloud computing BI software platform specializing in data visualization and dashboards. It is the fastest and easiest way to get insights from data. It offers the possibilty to create charts, reports, and dashboards from business data [61].

d) Looker

Is a cloud platform dedicated to data visualization. As a BI platform, Looker integrates several features, including many options dedicated to data visualization including the use of reporting tools, the creation of dashboards, and multicloud storage. Coders can use the LookML language to program visualization parameters.

6. Data Visualization challenges An interpretation is necessary before using quantitative data. Data visualizations combine the meaning of unprocessed data into meaningful conclusions. When designers put eye-catching visuals ahead of accuracy, the result is misleading visualizations. In order to convey data in an ethical manner, designers need to steer clear of common data visualization errors.

A. Algorithms and data inputs are susceptible to human error

Because human inputs are fallible, data visualization can only be as good as the people who provide it. Professionals who are unaware of the variations in applications may employ specific algorithms that emphasize specific information while ignoring other information. They might use a specific technique as a one-size-fits-all method for visualizing data, which can result in concepts being misrepresented. Analysts must take into account what makes each use case distinct.

TABLE VI. Top 15 JavaScript Visualization Libraries.

JavaScript Library	Description
D3JS	Best for document modification driven by data.
Charts.js	Best for plotting logarithmic, date, time, or custom scales on sparse and complex datasets.
FusionCharts	For needs in data visualization and charting for web and enterprise applications, it works best.
Taucharts	Ideal for groups creating intricate data displays.
Two.js	Best for a 2-D shape rendering open-source library.
Pts.js	Most effective for assembling items as you view them as points with a simple level of abstraction.
Raphael.js	Best for quickly and efficiently producing intricate drawings and graphics.
Anime.js	Best for producing potent user interface animation that is compatible with all of the main modern browsers.
ReCharts	Best for teams wanting to build charts for web applications based on React.
TradingVue.js	Best for creating complex charts, especially for stock and Forex trading applications on the web.
HighCharts	Best suited for teams needing a large library of charts to serve web and mobile platforms.
ChartKick	Best for producing simple charts using various libraries for programming languages, like Python, Ruby, JS, etc.
Pixi.js	Best for groups searching for JavaScript libraries to produce HTML5-based digital content.
Three.js	Best for creating 3-D visuals for applications on the web.
ZDog	The creation and rendering of 3-D pictures for canvas and SVG are not best for open-sourced software.

and employ a system that can help them achieve their particular objectives in order to minimize human error. Additionally, the use of machine learning and artificial intelligence can aid in lowering the requirement for human factors.

B. Data oversimplification

To make vast amounts of information easier to understand for viewers, visualizations condense them into simple graphs, scatter plots, and other visual aids. Because of this, some professionals have a

a tendency to oversimplify.

If they concentrate too much on the aesthetic appeal, they might overlook important details. As a result, viewers may draw incorrect inferences and conclusions from the imagery. In the end, this may result in poor decisions, which could be detrimental to businesses. By enlisting the aid of data analytics consulting firms, one can guarantee accurate representation of data while lowering the possibility of oversimplification.

C. Reliance on visualization is inevitable

Customers are depending more and more on data visualizations to understand their information. They make snap judgments based on visuals and aesthetics. It's a simple and efficient method of learning, and even as technology develops quickly, it should continue to be applicable. But in order for businesses to remain competitive, the trend of consumers depending too much on visualization forces them to employ analytical tools.

D. Data overload

When working with vast and complicated datasets, researchers frequently face the difficulty of data overload. This may lead to misunderstandings, annoyance, and mistakes in your analysis. How do you make sense of your data and prevent yourself from being overloaded with information? Here are some pointers to assist you in controlling the amount of data you analyze. In data visualization, information overload is a prevalent issue. When designers incorporate an increasing number of datasets into the display, data overload occurs. This makes the visualization difficult to understand as well as difficult to construct [62].

Lack of attention and prioritizing is a common cause of information overload. You should take a step back and assess each data set if you find that you've fallen victim to data overload.

7. DataViz' issues

A good visualization makes the data easy to understand so that viewers can rapidly draw conclusions. The inclusion of excessive information is one of the most frequent errors in data visualization. It is challenging for viewers to come up with takeaways because of this. Similar to how visualizations suffer from overuse of visual effects by designers.

• Why are most visualization designs ineffective? Because they are created for the incorrect audience, data visualizations frequently lack effectiveness. Poor end-user communication causes dashboards to lose some of their apparent usefulness. The audience for the dashboard must be identified before the data visualization design process can begin.

• What mistakes should be avoided in data visualization? Duplicate data, missed data, unmarked NA values, etc. are examples of common errors. For instance, the three pie chart sectors in this pie chart total up to 193%, which is illogical. Your final visualizations would be useless if the data contained such inaccuracies.

• Does data visualization require coding? You may quickly construct an interactive data visualization without writing any code. Spreadsheets and reports with a lot of text are insufficient for effectively presenting the data we found. This is why data visualization is necessary to show the data in a form that makes it easier for everyone to understand complex ideas.

• Why is misleading data bad? The audience could not notice the pertinent information if there is too much material offered or if it is irrelevant. It gets increasingly challenging to identify specific trends as there is more data provided at once. The public is frequently misled using sparse but pertinent material by over-informing them.

• Can the data be misleading? Due to the sampling technique employed to collect the data, the results may be deceptive. For instance, the type and size of the sample used in any statistic has a significant bearing on the results. Since many surveys and polls are directed at specific groups of people who give particular answers, the sample sizes tend to be tiny and skewed.

A. How to secure data privacy when sharing visualizations?

A strong technique to convey patterns, insights, and narratives from large, complicated data sets is through data visualization. However, how can you make sure that unauthorized users cannot access your visualizations and see private or sensitive data?

Individuals and organizations have the right and obligation to manage the collection, use, and sharing of their proprietary or personal data. This is known as data privacy. Data privacy is a business and reputational concern in addition to a legal and ethical one. You run the risk of facing legal repercussions, negative consumer feedback, and reputational harm if you neglect to secure data privacy.

Furthermore, you run the danger of losing the confidence of your stakeholders, partners, and data providers, which could have an impact on the availability and quality of your data.

Anonymizing data prior to visualizing it helps protect data privacy. In this procedure, any identifiers—such as names, addresses, phone numbers, or email addresses—that could be used to connect data to particular individuals or entities are either hidden or removed. A variety of techniques can be used to anonymize data, including substituting random or pseudonymous values for identifiers, aggregating or summarizing data at a finer level of detail, introducing noise or distortion to value data, choosing or sampling a subset of data, and switching or combining data values between records.

One practical method for guaranteeing data confidentiality is data encryption. This entails converting data into an encoded format that is only readable or accessible by individuals with the proper authorization and a decryption key or password. Data can be encrypted using a variety of techniques, including applying encryption algorithms or software, utilizing secure protocols or platforms, and incorporating encryption functions or libraries into encryption tools. visualization, as well as securing visualization outputs with a password or access control. All of these techniques can aid in preventing unwanted access to sensitive data.

B. AI's impact on data visualisation Work

Thanks to technology advancements, data visualization once restricted to simple graphs and tables now has a far more sophisticated aspect. As it positions itself as a major catalyst, artificial intelligence reveals new methods for accurately representing and interpreting the vast amount of information at our disposal. The nexus between Artificial Intelligence (AI) and visualization goes beyond a straightforward display; it is revolutionizing our comprehension and utilization of data [62].

In terms of theory and consulting, Chat Generative Pretrained Transformer (ChatGPT) still needed improvement because it was still prone to errors and lacked in-depth knowledge. In spite of this, it's still amazing how quickly and accurately it could respond in writing to my question. On the DataViz coding side was where ChatGPT truly excelled. We can ask ChatGPT to generate the code for a specific chart in multiple languages or libraries, and it will be as simple as that. In order to alter the code's appearance or functionality, we can also ask it to update certain parts of it. Even broken code can be fixed by ChatGPT, which can also tell you what went wrong [63], [64].

What effect will ChatGPT and similar AI tools have on DataViz work in the future, though, if they keep getting better? The way we retrieve information from the internet could already be revolutionized by the ChatGPT interface [65], [66]. To what extent, then, would AI tools disrupt the data visualization industry? In the following table VII, some educated guesses about what could take place in the future because of the AI.

8. Conclusions and limitations

With the development of smart technologies that generate astronomical amounts of data, data visualization becomes essential. Indeed, in order to be able to analyze your Big Data and make the best use of it in your business strategy, it is essential to be able to read it and list your business information in visual dashboards.

By classifying, segmenting and scripting data visually, a business can uncover previously inaccessible information at a glance. Data visualization therefore allows any organization to manage its activity more efficiently by adopting a data-driven and agile strategy. If data visualization was important a few years ago, it is crucial today. In the era of Big Data, it makes it possible to make sense of the billions of data that a company can collect every day and which, before this transformation process, are presented in separate lines and are therefore not easily exploitable. Data visualization is a very important task nowadays for data scientists. The main reason for recourse is decision-making.

An interpretable, relevant, and innovative visualization can lead to the right decision for a company knowing that this decision could be radical. Conventional visualization techniques cannot handle the enormous volume, variety, and velocity of data. To do this, several tools have emerged and are constantly evolving.

Therefore, modeling large data is a useful topic right now, among other things. Data modeling is actually a process that gives businesses access to a simple graphical user interface for finding, designing, visualizing, standardizing, and deploying high-quality data assets. Now, a sound data model acts as a guide for creating and implementing.

TABLE VII. Some educated guesses about what the future may bring negatively.

AI's DataViz issue	Description
Potential Job Losses	Several artists were upset when an AI-generated image won an art competition last year. With a description in natural language, DALL-E and its offspring can produce realistic-looking artwork and images.
Resource Websites Tools Redundancy	Perhaps as our methods of information acquisition change, AI tools will lessen the significance of webpages. Perhaps articles, blog posts, brochure websites, and even Wikipedia pages will become much less popular.
A Deluge of Mediocrity	This isn't unique to DataViz, but AI may have a very detrimental impact on the caliber of web content. There will probably be a spike in the quantity of AI-generated content available online as these tools enable faster and simpler content production. It's possible that online content created by AI will eventually surpass that created by humans.
Fake Data Visualisation	We might observe AI producing a ton of charts in the field of data visualization to further a cause. We learned how ChatGPT can create fake data in the previous post. Therefore, it is not out of the question that AI could create charts in large quantities using fictitious datasets. Even if a source is included at the bottom of the chart, it's possible that it was created entirely by AI to deceive those who were unwilling to look further.
Impact on Design Decisions	Over the last ten or so years, there has been a noticeable decline in attention spans. The growing appeal of Tiktok and YouTube shorts, particularly among younger audiences, makes this evident. It is unfortunate to say that if this trend keeps up, it will definitely affect how people look for and consume content that is data-focused.

databases that use higher-quality data sources to enhance application development and help users make wiser decisions [67]. So, we will be interested in big data modeling systems.

References

[1] K. T. Torphy, D. L. Brandon, A. J. Daly, K. A. Frank, C. Greenhow, S. Hu, and M. Rehm, "Social media, education, and digital democratization," Teachers College Record, vol. 122, no. 6, pp. 1–7, 2020.

[2] M.Walter, R. Lovett, B. Maher, B. Williamson, J. Prehn, G. BodkinAndrews, and V. Lee, "Indigenous

data sovereignty in the era of big data and open data," Australian Journal of Social Issues, vol. 56, no. 2, pp. 143–156, 2021.

[3] R. Mkhinini Gahar, A. Hidri, and M. Sassi Hidri, "Let's predict who will move to a new job," CoRR, vol. abs/2309.08333, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2309.08333

[4] M. Sassi Hidri, S. A. Alsaif, and A. Hidri, "Performance analysis of machine learning algorithms on networks intrusion detection," Int. J. Comput. Appl. Technol., vol. 70, no. 3/4, pp. 285–295, 2022. [Online]. Available: https://doi.org/10.1504/IJCAT.2022.10056028

[5] S. A. Alsaif, A. Hidri, and M. Sassi Hidri, "Towards inferring influential facebook users," Comput., vol. 10, no. 5, p. 62, 2021. [Online]. Available: https://doi.org/10.3390/computers10050062

[6] ——, "Stacking-based modelling for improved over-indebtedness predictions," Int. J. Comput. Appl. Technol., vol. 69, no. 3, pp. 273–281, 2022. [Online]. Available: https://doi.org/10.1504/IJCAT.2022.10052783

[7] M. A. Zoghlami, M. Sassi Hidri, and R. Ben Ayed, "Samplingbased consensus fuzzy clustering on big data," in 2016 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2016, Vancouver, BC, Canada, July 24-29, 2016. IEEE, 2016, pp. 1501–1508.

[8] R. Mkhinini Gahar, O. Arfaoui, M. Sassi Hidri, and N. Ben HadjAlouane, "An ontology-driven mapreduce framework for association rules mining in massive data," in Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 22nd International Conference KES-2018, Belgrade, Serbia, 3-5 September 2018, ser. Procedia Computer Science, R. J. Howlett, L. C. Jain, Z. Popovic, D. B. Popovic, S. N. Vukosavic, C. Toro, and Y. Hicks, Eds., vol. 126. Elsevier, 2018, pp. 224–233.

[9] G. Manogaran, D. Lopez, C. Thota, K. M. Abbas, S. Pyne, and R. Sundarasekar, "Big data analytics in healthcare internet of things," in Innovative healthcare systems for the 21st century. Springer, 2017, pp. 263–284.

[10] R. Mkhinini Gahar, O. Arfaoui, M. Sassi Hidri, and N. Ben HadjAlouane, "A distributed approach for high-dimensionality heterogeneous data reduction," IEEE Access, vol. 7, pp. 151006–151022, 2019.

[11] Z. M. Khalid, S. R. Zeebaree et al., "Big data analysis for data visualization: A review," International Journal of Science and Business, vol. 5, no. 2, pp. 64–75, 2021.

[12] A. Lensen, B. Xue, and M. Zhang, "Genetic programming for evolving a front of interpretable models for data visualization," IEEE transactions on cybernetics, vol. 51, no. 11, pp. 5468–5482, 2020.

[13] F. Skender, V. Manevska, I. Hristoski, and N. Rendevski, "Investigation of dataviz as a big data visualization tool," in International Symposium on Intelligent Manufacturing and Service Systems. Springer, 2023, pp. 469–478.

[14] A. M. S. R. N. Ibeh, "Dataviz design: A study of intentions," in 46 th CONFERENCE, 2023, p. 392.

[15] M. Friendly, "The golden age of statistical graphics," Statistical Science, pp. 502–535, 2008.

[16] T. Morita, "Reflection on the development of the tool kits of bertin's methods," Cartography and Geographic Information Science, vol. 46, no. 2, pp. 140–151, 2019.

[17] P. Organisciak, B. M. Schmidt, and J. S. Downie, "Giving shape to large digital libraries through exploratory data analysis," Journal of the Association for Information Science and Technology, vol. 73, no. 2, pp. 317–332, 2022.

[18] S. Jung, R. Xiao, O. Buruk, and J. Hamari, "Designing gaming wearables: From participatory design to concept creation," in Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction, 2021, pp. 1–14.

[19] L. Zhang, B. Vinodhini, and T. Maragatham, "Interactive iot data visualization for decision making in business intelligence," Arabian Journal for Science and Engineering, pp. 1–11, 2021.

[20] T. U. Nærland and M. Engebretsen, "Towards a critical understanding of data visualisation in

democracy: a deliberative systems approach," Information, Communication & Society, pp. 1–19, 2021.
[21] P. Raineri and F. Molinari, "Innovation in data visualisation for public policy making," in The Data Shake. Springer, Cham, 2021, pp. 47–59.
[22] J. B. Holbrook, "Open science, open access, and the democratization of knowledge," Issues in Science and Technology, vol. 35, no. 3, pp. 26–28, 2019.
[23] S. Sarica, B. Yan, G. Bulato, P. Jaipurkar, and J. Luo, "Data-driven network visualization for innovation and competitive intelligence," in Proceedings of the 52nd Hawaii International Conference on System Sciences, 2019.
[24] C. Burnett, G. Merchant, and I. Guest, "Destabilising data: The use of creative data visualisation to generate professional dialogue," British Educational Research Journal, vol. 47, no. 1, pp. 105–127, 2021.
[25] F. Deng, Z. Zhang, J. Zhang, and D. Zhang, "Building extraction from multiple images and lidar data," in Proceedings of the International conference on SAR and Multispectral Image Processing, vol. 6043, 2005, pp. 515–520.
[26] R. Grant, "Pretty persuasion: The advantages of data visualisation," Impact, vol. 2019, no. 2, pp. 19–23, 2019.
[27] S. Graessley, P. Suler, T. Kliestik, and E. Kicova, "Industrial big data analytics for cognitive internet of things: wireless sensor networks, smart computing algorithms, and machine learning techniques," Analysis and Metaphysics, vol. 18, pp. 23–29, 2019.
[28] O. Santolalla, "Dataviz," in Rock the Tech Stage. Springer, 2020, pp. 33–49.
[29] H. Oike, Y. Ogawa, and K. Oishi, "Simple and quick visualization of periodical data using microsoft excel," Methods and protocols, vol. 2, no. 4, p. 81, 2019.
[30] N. Akhtar, N. Tabassum, A. Perwej, and Y. Perwej, "Data analytics and visualization using tableau utilitarian for covid-19 (coronavirus)," Global Journal of Engineering and Technology Advances, vol. 3, pp. 28–50, 2020.
[31] S. Hiljazi and T. Curtis, "Developing an introductory class in business intelligence (bi) using ms excel powerpivot." Association Supporting Computer Users in Education, 2018.
[32] J. Dougherty and I. Ilyankou, Hands-On Data Visualization. " O'Reilly Media, Inc.", 2021.
[33] J. Doshi, A. Goradia, and D. Mistry, "A review of google data visualization tools," International Journal of Current Engineering and Technology, vol. 4, no. 5, pp. 3134–3138, 2014.
[34] J. A. S. Cardona and D. A. A. Garcia, "Evaluaci´ on y selecci´ on de herramientas de anal´ ıtica visual para su implementaci´on en una instituci´on de educaci´ on superior," Revista IngEam, vol. 4, no. 1, pp. 1–20, 2017.
[35] D. Orlovskyi, A. Kopp, and V. Kondratiev, "Using dashboards for the business processes status analysis," 2019.
[36] M. Minelli, M. Chambers, and A. Dhiraj, Big data, big analytics: emerging business intelligence and analytic trends for today's businesses. John Wiley & Sons, 2013, vol. 578.
[37] M. H. u. Rehman, V. Chang, A. Batool, and T. Y. Wah, "Big data reduction framework for value creation in sustainable enterprises," Int. J. Inf. Manag., vol. 36, no. 6, p. 917–928, dec 2016.
[38] H. Oliv´ er, "K´ ıs´erleti gy´art´ ashoz kapcsol´od´o adatvizualiz´aci´os fejleszt´es a purt´ar rendszerben," Multidiszciplin´aris Tudom´ anyok, vol. 10, no. 4, pp. 238–252, 2020.
[39] T. P. Atwood and R. Reznik-Zellen, "Using the visualization software evaluation rubric to explore six freely available visualization applications," Journal of eScience Librarianship, vol. 7, no. 1, 2018.
[40] K. Wilson, "Microsoft office 365," in Using office 365. Springer, 2014, pp. 1–14.
[41] H. Kennedy and W. Allen, "Data visualisation as an emerging tool for online research," The Sage handbook of online research methods, pp. 307–326, 2016.

[42] M. Kaufmann, "Big data management canvas: A reference model for value creation from data," Big Data and Cognitive Computing, vol. 3, no. 1, 2019.

[43] N. C. Viorel and N. Lucia, "Analysis of information on tourism in the european union using the power bi business analysis service." Agricultural Management/Lucrari Stiintifice Seria I, Management Agricol, vol. 21, no. 1, 2019.

[44] S. A. Fahad and A. E. Yahya, "Big data visualization: allotting by R and python with GUI tools," in Proceedings of the IEEE International Conference on Smart Computing and Electronic Enterprise, 2018, pp. 1–8.

[45] S. N. Pattanaik and R. P. Wiegand, "Data visualization," Handbook of Human Factors and Ergonomics, pp. 893–946, 2021.

[46] J. Richardson, R. Sallam, K. Schlegel, A. Kronz, and J. Sun, "Magic quadrant for analytics and business intelligence platforms," Gartner ID G00386610, 2020.

[47] S. Khan, "Data visualization to explore the countries dataset for pattern creation," International Journal of Online & Biomedical Engineering, vol. 17, no. 13, 2021.

[48] G. Srivastava and R. Venkataraman, "A review of the state of the art in business intelligence software," Enterprise Information Systems, vol. 16, no. 1, pp. 1–28, 2022.

[49] A. M. Amer and M. M. El-Hadi, "Tableau big data visualization tool in the higher education institutions for sustainable development goals," International Journal of Computer Science and Mobile Computing, 2019.

[50] I. Viljanen, "Improving solutions for analytics services in a midsized insurance company," 2020.

[51] V. Vashisht and P. Dharia, "Integrating chatbot application with qlik sense business intelligence (BI) tool using natural language processing (NLP)," in Micro-Electronics and Telecommunication Engineering. Springer, 2020, pp. 683–692.

[52] G. Sedrakyan, J. Malmberg, K. Verbert, S. J¨ arvel¨a, and P. A. Kirschner, "Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation," Computers in Human Behavior, vol. 107, p. 105512, 2020.

[53] M. Islam and S. Jin, "An overview of data visualization," inxProceedings of the International Conference on Information Science and Communications Technologies (ICISCT), 2019, pp. 1–7.

[54] T. Usova and R. Laws, "Teaching a one-credit course on data literacy and data visualisation." Journal of Information Literacy, vol. 15, no. 1, 2021.

[55] S.K.Peddoju and H. Upadhyay, "Evaluation of iot data visualization tools and techniques," in Data visualization. Springer, 2020, pp. 115–139.

[56] M.Mauri, T. Elli, G. Caviglia, G. Uboldi, and M. Azzi, "Rawgraphs: a visualisation platform to create open outputs," in Proceedings of the 12th biannual conference on Italian SIGCHI, 2017, pp. 1–5.

[57] F. B. Viegas, M. Wattenberg, and J. Feinberg, "Participatory visualization with wordle," IEEE transactions on visualization and computer graphics, vol. 15, no. 6, pp. 1137–1144, 2009.

[58] A. Weiner and K. Lorber, "Infographics: A methodology for student research presentations and other academic projects," in Proceedings of the International Conference on Society for Information Technology & Teacher Education. Association for the Advancement of Computing in Education (AACE), 2021, pp. 649–652.

[59] M.Serik, G. Nurbekova, and M. Mukhambetova, "Optimal organisation of a big data training course: big data processing with bigquery and setting up a dataproc hadoop framework," World Trans. on Engng. and Technol. Educ, vol. 19, no. 4, pp. 417–422, 2021.

[60] B. Arruabarrena, "L'expert en dataviz, un m´etier en transition," I2DInformation, donn´ ees documents, vol. 54, no. 3, pp. 7–8, 2017.

[61] A. M. Pedersen and C. Bossen, "Data work in healthcare: An ethnography of a bi unit," in

Proceedings of the 8th International Conference on Infrastructures in Healthcare. European Society for Socially Embedded Technologies (EUSSET), 2021.

[62] R. Rozi´ c, R. Sli˘ skovi´c, and M. Rosi´c, "Artificial intelligence for knowledge visualization: An overview," in International Conference on Digital Transformation in Education and Artificial Intelligence Application. Springer, 2023, pp. 118–131.

[63] B. Kovalerchuk, K. Nazemi, R. Andonie, N. Datia, and E. Banissi, Integrating Artificial Intelligence and Visualization for Visual Knowledge Discovery. Springer, 2022.

[64] C. Conner, J. Samuel, M. Garvey, Y. Samuel, and A. Kretinin, "Conceptual frameworks for big data visualization: Discussion of models, methods, and artificial intelligence for graphical representations of data," in Handbook of Research for Big Data. Apple Academic Press, 2022, pp. 197–234.

[65] A. Lavanya, S. Sindhuja, L. Gaurav, and W. Ali, "A comprehensive review of data visualization tools: Features, strengths, and weaknesses," 2023.

[66] N. W. Kim, G. Myers, and B. Bach, "How good is chatgpt in giving advice on your visualization design?" arXiv preprint arXiv:2310.09617, 2023.

[67] R. Mkhinini Gahar, O. Arfaoui, and M. Sassi Hidri, "Towards big.

Rania Mkhinini Gahar was born in M'Saken, Sousse, Tunisia. She received the Engineering degree from the Higher Institute of Applied Sciences and Technology of Sousse, Tunisia, in 2013, the master's degree and the Ph.D degree from the National Engineering School of Tunis, Tunisia, respectively in 2015 and 2021. She is a member of the OASIS Research Lab with the National School of Engineers of Tunis. Her research interests include big data analytics, as well as data science and statistical machine learning.

Olfa Arfaoui was born in Bou Arada, Seliana, Tunisia. She received the Engineering degree in computer science from the National School of Engineering of Tunis (ENIT), Tunis, EL Manar University, Tunisia, in 2007, and the Ph.D. degree from ENIT, Tunisia, in 2014. She is currently an Assistant Professor with the University of Carthage, Tunisia. Her research activities focus on clustering, XML native databases, flexible querying, big data analytics, as well as data science.

Minyar Sassi Hidri was born in Nabeul, Tunisia. She received the degree in computer science engineering and the Ph.D. degree from the National Engineering School of Tunis (ENIT), Tunis El Manar University, Tunisia, in 2003 and 2007, respectively. She obtained the Qualification to lead researches in computer sciences from Tunis El Manar University, Tunisia, in June 2014. She is currently an Associate Professor with ENIT, Tunisia, and an Assistant Professor with the Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia, since September 2017. Her experience in teaching, in computer science and information systems is around 20 years. Her research interests include combinatorial aspects in Big Data analytics, machine learning, deep learning, and text mining, with over 65 publications. She is also a member of the steering committee of many international conferences and a reviewer of impacted journals.

Instructions for Authors

Essentials for Publishing in this Journal

1 Submitted articles should not have been previously published or be currently under consideration for publication elsewhere.

2 Conference papers may only be submitted if the paper has been completely re-written (taken to mean more than 50%) and the author has cleared any necessary permission with the copyright owner if it has been previously copyrighted.

3 All our articles are refereed through a double-blind process.

4 All authors must declare they have read and agreed to the content of the submitted article and must sign a declaration correspond to the originality of the article.

Submission Process

All articles for this journal must be submitted using our online submissions system. http://enrichedpub.com/ . Please use the Submit Your Article link in the Author Service area.

Manuscript Guidelines

The instructions to authors about the article preparation for publication in the Manuscripts are submitted online, through the e-Ur (Electronic editing) system, developed by **Enriched Publications Pvt. Ltd**. The article should contain the abstract with keywords, introduction, body, conclusion, references and the summary in English language (without heading and subheading enumeration). The article length should not exceed 16 pages of A4 paper format.

Title

The title should be informative. It is in both Journal's and author's best interest to use terms suitable. For indexing and word search. If there are no such terms in the title, the author is strongly advised to add a subtitle. The title should be given in English as well. The titles precede the abstract and the summary in an appropriate language.

Letterhead Title

The letterhead title is given at a top of each page for easier identification of article copies in an Electronic form in particular. It contains the author's surname and first name initial .article title, journal title and collation (year, volume, and issue, first and last page). The journal and article titles can be given in a shortened form.

Author's Name

Full name(s) of author(s) should be used. It is advisable to give the middle initial. Names are given in their original form.

Contact Details

The postal address or the e-mail address of the author (usually of the first one if there are more Authors) is given in the footnote at the bottom of the first page.

Type of Articles

Classification of articles is a duty of the editorial staff and is of special importance. Referees and the members of the editorial staff, or section editors, can propose a category, but the editor-in-chief has the sole responsibility for their classification. Journal articles are classified as follows:

Scientific articles:

1. Original scientific paper (giving the previously unpublished results of the author's own research based on management methods).

2. Survey paper (giving an original, detailed and critical view of a research problem or an area to which the author has made a contribution visible through his self-citation);

3. Short or preliminary communication (original management paper of full format but of a smaller extent or of a preliminary character);

4. Scientific critique or forum (discussion on a particular scientific topic, based exclusively on management argumentation) and commentaries. Exceptionally, in particular areas, a scientific paper in the Journal can be in a form of a monograph or a critical edition of scientific data (historical, archival, lexicographic, bibliographic, data survey, etc.) which were unknown or hardly accessible for scientific research.

Professional articles:

1. Professional paper (contribution offering experience useful for improvement of professional practice but not necessarily based on scientific methods);
2. Informative contribution (editorial, commentary, etc.);
3. Review (of a book, software, case study, scientific event, etc.)

Language

The article should be in English. The grammar and style of the article should be of good quality. The systematized text should be without abbreviations (except standard ones). All measurements must be in SI units. The sequence of formulae is denoted in Arabic numerals in parentheses on the right-hand side.

Abstract and Summary

An abstract is a concise informative presentation of the article content for fast and accurate Evaluation of its relevance. It is both in the Editorial Office's and the author's best interest for an abstract to contain terms often used for indexing and article search. The abstract describes the purpose of the study and the methods, outlines the findings and state the conclusions. A 100- to 250-Word abstract should be placed between the title and the keywords with the body text to follow. Besides an abstract are advised to have a summary in English, at the end of the article, after the Reference list. The summary should be structured and long up to 1/10 of the article length (it is more extensive than the abstract).

Keywords

Keywords are terms or phrases showing adequately the article content for indexing and search purposes. They should be allocated heaving in mind widely accepted international sources (index, dictionary or thesaurus), such as the Web of Science keyword list for science in general. The higher their usage frequency is the better. Up to 10 keywords immediately follow the abstract and the summary, in respective languages.

Acknowledgements

The name and the number of the project or programmed within which the article was realized is given in a separate note at the bottom of the first page together with the name of the institution which financially supported the project or programmed.

Tables and Illustrations

All the captions should be in the original language as well as in English, together with the texts in illustrations if possible. Tables are typed in the same style as the text and are denoted by numerals at the top. Photographs and drawings, placed appropriately in the text, should be clear, precise and suitable for reproduction. Drawings should be created in Word or Corel.

Citation in the Text

Citation in the text must be uniform. When citing references in the text, use the reference number set in square brackets from the Reference list at the end of the article.

Footnotes

Footnotes are given at the bottom of the page with the text they refer to. They can contain less relevant details, additional explanations or used sources (e.g. scientific material, manuals). They cannot replace the cited literature.
The article should be accompanied with a cover letter with the information about the author(s): surname, middle initial, first name, and citizen personal number, rank, title, e-mail address, and affiliation address, home address including municipality, phone number in the office and at home (or a mobile phone number). The cover letter should state the type of the article and tell which illustrations are original and which are not.

Note